Drivers of variability in the structure and function of marine microbial communities: from cell physiology to the global environment

Nicholas Bock

Submitted in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy
under the Executive Committee
of the Graduate School of Arts and Sciences

COLUMBIA UNIVERSITY

2021

Abstract

Drivers of variability in the structure and function of marine microbial communities: from cell physiology to the global environment

Nicholas Bock

Marine microorganisms are a key vector in global carbon cycling, supporting an annual flux of 5 – 12 gigatons of carbon to the ocean interior via the biological carbon pump. While methodological advances over the last half century have greatly advanced our understanding of the factors influencing variability in this flux, the contributions of individual components in the microbial food web remain poorly resolved. Utilizing a combination of laboratory, field and remote sensing studies, this dissertation addresses several different aspects of this challenge. In the second chapter, unsupervised learning methods are applied to a global bio-optical data set from biogeochemical Argo floats to identify six oceanic biomes characterized by distinct seasonal trends in vertical phytoplankton distributions. This study demonstrated the great potential for using data from autonomous profiling floats to generalize seasonal trends in vertical phytoplankton distributions across vast regions of the global ocean, while also providing new insight on the hydrological and biogeochemical drivers of this variability. The third chapter reports the development of a novel method for the direct measurement of chlorophyll *a* attributable to individual phytoplankton groups in natural samples via cell sorting by flow cytometry. Critically, this approach makes it possible to evaluate phytoplankton community structure in terms of a parameter measured by autonomous platforms, while simultaneously quantifying sources of variability not captured by existing methods. The fourth chapter investigates the environmental drivers of phytoplankton distributions within the Western Tropical South Pacific, providing a case study for the biogeographical provinces identified in

chapter 2 while also investigating how biogeochemical gradients influence linkages between heterotrophic groups central to carbon cycling within the microbial food web. Chapter five reports series of experiments investigating cell physiology as a driver of predator-prey interactions between heterotrophic bacteria and algal phagomixotrophs—eukaryotic algae that supplement requirements for carbon and/or nutrients by ingesting smaller cells. By validating the predictions of a gene-based model of algal trophic modes, the results from these experiments point toward the potential widespread occurrence of phagomixotrophy amongst green algae, while highlighting potential sources of bias in field and laboratory studies of bacterivory. With global climate change expected to produce rapid changes in ocean circulation and biogeochemistry, the urgency of understanding the role of marine microbes in global biogeochemical cycling has never been greater. This dissertation represents an advance in this larger goal, providing an expanded framework for the broad distribution of microbial communities in addition to novel insight into the environmental and physiological drivers of microbial community structure from the global to cellular scale.

Table of Contents

List of Tables

List of Figures

Acknowledgments

I cannot sufficiently put into words how much I have enjoyed having the opportunity to study and conduct research at Lamont Doherty Earth Observatory (LDEO) over the last five years. Between the wonderful and supportive research community at Lamont, the talented students, faculty, and staff that I have been able to work with, and all the long bike rides through the Palisades, it has been an amazing experience.

Above all, I am deeply indebted to my research advisor, Dr. Solange Duhamel, for supporting my application to the Department of Earth and Environmental Sciences (DEES) PhD program and for all of her support and guidance in completing the research described in this dissertation. Aside from making herself virtually always available to brainstorm an idea, troubleshoot an instrument or provide insight on a figure, the latitude that Dr. Duhamel has permitted me in exploring the subfields of marine microbial ecology has made the last five years continually challenging and rewarding in equal measure. In her dedication to research, her sharp insight, and her limitless dedication to her mentees, the is truly an inspiration. I am fortunate to have had the opportunity to work with her.

I also am grateful for the support of my advisory committee: Dr. O. Roger Anderson, Dr. Eunsoo Kim, Dr. Andrew Juhl. I thank Dr. Anderson for his limitless generosity, support, and encouragement over the last five years, and for providing an opportunity to conduct research at LDEO while I was still a student at Teachers College. I thank Dr. Kim for welcoming me into her research group at the American Museum of Natural History (AMNH), her guidance and training in epifluorescent microscopy, and her endless optimism while waiting for feeding signals. I thank Dr. Juhl for always taking the time to talk through the details of instrumentation, analyses, lab administration, and virtually any other topic imaginable. I also thank Dr. Juhl for

fostering a sense of community amongst the biological oceanography faculty and staff through all the upheavals of the past few years. Finally, I thank Dr. Thierry Moutin at the Insitut Méditerranéen d'Oceanologie (MIO) for agreeing to lend his expertise as a member of my dissertation committee.

I would also like to thank my collaborators on the different chapters of my dissertation Dr. Hervé Claustre and Dr. Marin Cornec at Laboratoire d'Océanographie de Villefranche for generously providing their encouragement and expertise in the development of Chapter 2. Dr. Ajit Subramaniam for his guidance in the research supporting Chapter 3, and for constantly challenging me to interrogate the assumptions, motivations, and feasibility of my research objectives. Dr. France Van Wambecke, also at MIO, for all her generous guidance and support in completing the analyses for Chapter 4. I would also like to thank Dr. Sophie Charvet, now at AMNH, for being an invaluable mentor and friend over the last four years, for always rallying to catch the first shuttle up and last shuttle back, and for keeping up morale during feeding experiment after feeding experiment.

I also owe a debt of gratitude to the support and guidance provided by the all the additional co-authors who have lent their insight and expertise. Dr. Joseph Montoya for the leadership provided aboard the R/V Endeavor, and for his support in the development of Chapter 3. Moïra Dion at Université Laval for processing the innumerable cytometry samples collected during the OUTPACE cruise and for her perspective on the corresponding analyses. Dr. John Burns, now at Bigelow Laboratory for Ocean Science, for his mentorship and support at both AMNH and Lamont, and for his unfailing encouragement during the experiments supporting Chapter 5. I also thank Dr. Andrey Rosenberg at Israel Institute of Technology for his guidance in developing the manuscript that resulted from the work described in Chapter 5.

I am also grateful to the administrators of the DEES PhD program: Kaleigh Matthews, Carol Mountain, Sally Odland, Yasmin Yabyabin, and Monica Hinojosa. I would also like to give a shout out to the other students in the DEES PhD program, especially Elise Myers, Rebecca Trinh, and Kailani Acosta, who have been a constant source of support and optimism in navigating research, coursework, and the various challenges and setbacks encountered during the PhD experience. I would be remiss not to thank Benjamin Ramcharitar for his endless responsiveness and flexibility in completing experiments and analyses, and the captain and crew of the R/V Endeavor during EN620 and EN642 for their patience, comradery, and support in the field.

Finally, I would like to thank my family. Andrea for all her encouragement and support, for reading countless drafts, for attentively listening to practice rounds of even the most half-formed presentations, and for letting me vent about CellTracker over innumerable dinners. I thank my parents for their encouragement and guidance both during the PhD program and for being such wonderful sounding boards as I consider what comes next. I also thank Andrea's parents, Dani and Henry Jakubowski, for celebrating every success along the way, and for all their countless hours of help taking care of Frances over the last year.

To Andrea and Frances

Chapter 1: Introduction

Marine environments represent a key term in global carbon cycling, taking up an estimated 92.4 gigatons of carbon annually: a flux driven by the solubility of CO_2 in seawater and by the assimilation of atmospheric dissolved inorganic carbon (DIC) by marine organisms (Schlesinger and Bernhardt 2013). While DIC is readily exchanged across the atmosphere-sea interface to maintain chemical equilibrium, biological carbon fixation supports a flux thought to deliver 5 – 12 gigatons of carbon to the ocean interior annually via the biological carbon pump (Laws et al. 2000; Henson et al. 2011). With marine microbes dominating primary production in the global ocean, (Ducklow 1983; Field et al. 1998), the strength and functioning of the microbial carbon pump is primarily a function of the distribution and physiology of these organisms. Therefore, understanding the factors that shape the distribution and function of microbial communities is crucial to understanding marine carbon fluxes and long-term carbon sequestration (Guidi et al. 2009).

While the widespread distribution of marine microbes has been recognized for centuries (Karl and Proctor 2007), the current understanding of their role in carbon cycling has primarily been shaped by methodological advances made over the last seventy-five years, including the use of ^{14}C to measure phytoplankton primary production (Steemann Nielsen 1952); the utilization of membrane filters for evaluating planktonic biomass and physiology across discrete size classes (Fleischer et al. 1964); and the development of instruments for assessing *in situ* distributions of algal particles via chlorophyll fluorescence (Holm-hansen et al. 1965).

Over the last 40 years in particular, improved methodologies in the optical characterization of microorganisms have transformed conceptions of the abundance and diversity of marine microbes. The development of epifluorescent microscopy methodologies made evident

the enormous abundance of both heterotrophic (Hobbie et al. 1977) and autotrophic (Waterbury et al. 1979) prokaryotes (hereafter bacteria) in marine environments. Subsequent studies (Azam et al. 1983; Ducklow et al. 1986) upended the perceived dominance of the algae to zooplankton to fishes food chain, revealing that the majority of carbon cycling within marine ecosystems is carried out by organisms within the "microbial loop"—wherein inputs of dissolved organic carbon (DOC) from photosynthetic organisms and heterotrophic protists support populations of heterotrophic bacteria.

At the same time, advancements in high-throughput methods such as flow cytometry have allowed for rapid determination of planktonic assemblages and abundance in field samples and for the segregation of cells for the study of group specific physiological processes (Yentsch et al. 1983). Flow cytometric characterization of phytoplankton communities also led to the discovery of *Prochlorococcus*, marine cyanobacteria ubiquitous in open ocean environments (Chisholm et al. 1988), and the discrimination of physiologically distinct subpopulations of heterotrophic bacteria (Li 1995; Vaqué et al. 2001). Further, the application of bioinformatics to phytoplankton genomic and transcriptomic sequence data has provided new insight on phytoplankton diversity and physiological responses to environmental variability.

Parallel to these developments in field and laboratory approaches, the deployment of optical sensors aboard remote sensing platforms has made it possible to study the sea surface at the global scale. Starting with the Color Zone Color Scanner, data from these platforms has been leveraged to develop algorithms for estimating surface concentrations of Chlorophyll *a* (Chl*a*; Smith and Baker 1982) and optical backscatter coefficients (b_{bp}), commonly used as a proxy for concentrations of particulate organic carbon (POC) in seawater (Loisel and Stramski 2000). Additionally, within the last decade, the widespread deployment of autonomous floats in the

biogeochemical Argo (BGC-Argo) network has provided subsurface measurements of these bio-optical parameters at spatiotemporal frequencies far exceeding what might be achievable based on ship data alone (Claustre et al. 2020).

Refinement of methodologies for studying microbial communities at both the scale of individual seawater samples and the scale of the global ocean has supported the development of approaches for estimating planktonic physiology based on synoptic bio-optical measurements. For example, global primary production models have been developed based on remote sensed Chl*a* in combination with modeled photosynthetic efficiency parameters (Behrenfeld and Falkowski 1997), ratios of Chl*a* and b_{bp} (Westberry et al. 2008), or as a product of photosynthetic active radiation (PAR) and modeled photosynthetic efficiency (Silsbe et al. 2016). These models have in turn made it possible to constrain the magnitude of carbon fluxes attributable to phytoplankton in global biogeochemical cycling (Field et al. 1998) and to evaluate phytoplankton primary production responses to interannual climate variability (Behrenfeld et al. 2006a; Rousseaux and Gregg 2015).

However, because these and related approaches are based on bulk estimates of planktonic biomass or physiology (e.g., remote sensed Chl*a* or b_{bp}), they necessarily provide estimates that reflect the physiology of the total phytoplankton community. They cannot, for instance, account for variability in primary production due to variability in the taxonomic composition of phytoplankton communities (Zubkov 2014), or to differential responses to resource limitation across individual taxa or size classes (Edwards et al. 2015). Moreover, by assuming all algal biomass represents autotrophic primary production, they cannot account for complexities introduced by combinations of trophic strategies, for instance phagomixotrophy: the supplementation of autotrophic metabolism via the phagocytotic uptake of organic particles,

including other cells (Ward and Follows 2016). Therefore, a central challenge in oceanography is to extend existing methodologies and frameworks to support inferences on physiological processes at the level of individual phytoplankton functional groups or size classes. This dissertation focuses on several different aspects of this goal.

In chapter 2, I applied empirical orthogonal function (EOF) analysis to a global data set of Chl*a* and b_{bp} time series from BGC-Argo floats, identifying six broad plankton biomes in the global ocean. Uniquely, the use of EOF analysis makes it possible to define these biomes without any prior assumptions regarding the biogeochemical, or hydrological parameters to be used as classification criteria. Moreover, by utilizing float data, this analysis accounts for subsurface planktonic distributions not visible based on surface measurements alone. The variability in bio-optical parameters associated with each biome integrates both bottom-up and top-down controls on microbial distributions, in addition to physiological responses to resource limitation across horizontal and vertical gradients. Chapters 3 – 5 investigate these sources of variability in detail.

Chapter 3 reports a flow cytometry method for deconvolving bulk measurements of phytoplankton chlorophyll concentration, such as those provided by BGC-Argo floats, to determine Chl*a* attributable to individual phytoplankton groups. This method uniquely 1) makes it possible to characterize group-specific changes in cellular Chl*a* concentration with depth, and 2) allows for the quantification of the fraction of bulk Chl*a* attributable to specific phytoplankton groups. I provide a case study of this approach applied to samples from the Gulf of Mexico, demonstrating the method to converge with existing methods for estimates of Chl*a* attributable to *Prochlorococcus* populations, while also making possible for the first time direct estimates of Chl*a* attributable to *Synechococcus* and picophytoeukaryotes (eukaryotic phytoplankton between 1 and 3 μm in diameter).

Chapter 4 investigates variability in planktonic community structure and function across a biogeochemical gradient spanning mesotrophic to oligotrophic conditions in the Western Tropical South Pacific. In this analysis, I characterized distributions of phytoplankton biomass and applied an empirical model to evaluate coupling between heterotrophic protists and heterotrophic bacteria in the microbial loop. My results highlighted the importance of picophytoeukaryotes to overall phytoplankton biomass in oligotrophic environments, while also capturing a decoupling of heterotrophic populations under nutrient depleted conditions, implying increased competition for bacterial prey under limitation of dissolved nutrients (Bock et al. 2018).

To measure such interactions directly, in chapter 5 I developed a novel method for assessing bacterivory via microscopy and flow cytometry using live fluorescently labeled bacteria. Applying this method to laboratory cultures of pico- and nano-sized phytoplankton, I identified possible sources of bias in the determination of bacterivory when following existing methods and reported the positive identification of bacterivory in five strains of *Prasinophyceae* not previously known to be phagomixotrophic (Bock et al. 2021). Moreover, by validating predictions provided by a gene-based model of algal trophic modes, these results additionally underlined the likelihood of widespread phagomixotrophy amongst green algae.

Chapter 2: Biogeographical classification of the global ocean from BGC-Argo floats

2.1. Abstract

Biogeographical classifications of the global ocean generalize spatiotemporal trends in distributions of species or biomass distributions across discrete ocean biomes or provinces. These classifications are generally based on a combination of remote-sensed proxies of phytoplankton biomass and global climatologies of biogeochemical or physical parameters. However, these approaches are limited in their capacity to account for subsurface variability in these parameters. The deployment of autonomous profiling floats in the biogeochemical Argo (BGC-Argo) network over the last 10 years has greatly increased global coverage of subsurface measurements of bio-optical proxies for phytoplankton biomass and physiology. In this study, we used empirical orthogonal function analysis to identify the main sources of variability in a data set of 422 time series of vertical profiles chlorophyll *a* fluorescence and optical backscatter coefficients. Applying cluster analysis to these results, we identified six phytoplankton biomes within the global ocean: two high-latitude biomes capturing summer bloom dynamics in the North Atlantic and Southern Ocean and four mid to low-latitude biomes characterized by variability in the depth and frequency of DCM formation. We report the distribution of these biomes and associated trends in biogeochemical and hydrological parameters. Our results demonstrate the nitracline depth to be well correlated with the DCM across low-latitude environments, while additionally highlighting light level at the nitracline depth to be a key determinant of biomass accumulation in both the euphotic zone and upper mesopelagic. In addition to partitioning seasonal variability in vertical phytoplankton distributions at the global scale, our results provide a potentially novel biogeographical classification of the global ocean.

2.2. Introduction

Biogeographical partitions of marine environments aim to identify ecologically similar ocean regions at the basin and/or global scale. Such partitions allow for the extrapolation of sparse ship-based measurements, in turn making it possible to generalize relationships between environmental variables and species distributions (Longhurst 2007), tune global bio-optical algorithms to specific regions or water types (Moore et al. 2009; Shi et al. 2009), and constrain expected variability in global models of biological processes (Platt and Sathyendranath 1988; Behrenfeld et al. 2005).

Because the majority of ocean primary production and biogeochemical cycling is carried out by microbial communities (Field et al. 1998), there is a particular interest in delineating regions characterized by similar trends in plankton dynamics. As such, biogeochemical classifications are often based on criteria presumed to reflect or at least correlate with the distribution or physiology of microbial communities. For example, Platt and Sathyendranath (1988) identified nine ocean regions based on bathymetric elevation and latitude. Longhurst subsequently identified four general ocean biomes based on seasonal trends in vertical mixing, using these biomes as the basis for identifying 51 provinces based on metrics of stratification, estimated photic zone depths and surface nutrient concentrations (Longhurst et al. 1995; Longhurst 2007). Other strategies have leveraged global measurements of remote sensed parameters to establish classifications, including sea surface temperature (SST), photosynthetically active radiation (PAR), and estimated concentration of Chlorophyll *a* ([Chl*a*]) at both regional (Hardman-Mountford and McGlade 2002; D'Ortenzio and Ribera d'Alcalà 2008) and global scales (Moore et al. 2001; Wilson and Coles 2005; Hardman-Mountford et al. 2008; Mustapha et al. 2014).

While these and other approaches (see Krug et al. 2017 for a review) have been successful at defining marine provinces based on the criteria selected, their necessary reliance on remote sensed parameters and/or annually averaged climatologies limits their ability to account for subsurface variability in their classifications. In the last 10 years, however, the widespread deployment of Biogeochemical Argo (BGC-Argo) floats has considerably relieved this constraint, providing high frequency measurements of physical, biogeochemical and bio-optical parameters in the horizontal, vertical and temporal dimensions over long time periods. These data have made it possible to characterize subsurface phytoplankton distributions and their drivers at both the basin (e.g., Mignot et al. 2014a; Lacour et al. 2017; Barbieux et al. 2019), and the global scale (Cornec et al. 2021).

The global coherence of measurements from BGC-Argo floats makes these data particularly amenable to unsupervised machine learning methods—analysis methods including principal components analysis (PCA), empirical orthogonal function analysis (EOF), and various clustering techniques—that make it possible to identify subgroups within a data set based on the data structure alone—that is, without requiring grouping criteria *a priori* (James et al. 2013). Previous studies have used such methods, namely EOF paired with cluster analysis, to characterize the main sources of variability in satellite measurements of bio-optical and physical parameters (Sharples et al. 2001; Hardman-Mountford et al. 2008). However, observation gaps in ship-based field measurements have made it impossible to characterize variability in subsurface measurements using similar approaches.

In this study we apply EOF analysis to 1-year time series of bio-optical parameters measured by BGC-Argo floats: *in vivo* fluorescence (hereafter [Chl*a*]) and particle backscattering coefficients (hereafter b_{bp}). At a first order, [Chl*a*] provides a coarse measure of

algal biomass, although it can also be significantly influenced by phytoplankton physiology, especially across vertical gradients (Dubinsky and Stambler 2009). Particle backscatter coefficients provide an overall proxy for particulate organic carbon (POC), especially within the 0.2 – 20 μm size range (Stramski et al. 2008). Incorporation of both [Chl*a*] and b_{bp} in EOF analysis makes it possible to compare individual time series based on variability in both phytoplankton biomass/physiology and overall POC distributions.

Based on this analysis, we address the following objectives: 1) to describe the dominant spatiotemporal trends in vertical phytoplankton and suspended particle distributions within the global ocean; 2) to identify the environmental drivers associated with this variably; 3) to evaluate how this variability in the upper ocean drives biomass accumulation within the mesopelagic zone. Our results provide the first ever biogeographical classification of the global ocean that relies solely on bio-optical parameters measured by BGC-Argo floats and yield novel insights on the drivers of phytoplankton dynamics at the global scale.

2.3. Methods

2.3.1. Construction of bio-optical time series

This study relies on data acquisition from 205 BGC-Argo floats sampling various regions of the global ocean (Argo 2000). The network of those autonomous platforms, possible thanks to the miniaturization of sensors and an international coordination of deployment (Chai et al. 2020), provides high spatial and temporal vertical profiles of physical and biogeochemical parameters (Roemmich et al. 2019; Claustre et al. 2020). Besides temperature and salinity, these measurements include vertical profiles of [Chl*a*] fluorescence, and up to four additional core BGC variables (e.g., PAR, dissolved oxygen, nitrate concentration and pH). BGC-Argo float data are accessible through the Coriolis database (ftp://ftp.ifremer.fr/ifremer/argo).

Hydrological data were collected by the SBE 41 Seabird CTD sensors were processed and quality-controlled following Argo protocols as described by Wong et al. (2020). The floats were equipped with Seabird-Wetlabs sensors of three types (FLBB, ECO-Triplet, or MCOMS) that include a Chl*a* fluorometer (excitation at 470 nm; emission at 695 nm), and a backscattering sensor at 700 nm. [Chl*a*] and b_{bp} were retrieved from those measurements and quality-controlled (only data with good and probably good QC flags were used) following BGC-Argo protocols (Schmechtig et al. 2015, 2018a, b). Individual [Chl*a*] and b_{bp} profiles were smoothed prior to EOF analysis in order to denoise data and remove spikes. Following Cornec et al. (2021), a 5-point rolling median filter was applied to [Chl*a*] profiles. For b_{bp} profiles, a 7-point rolling median filter was applied, followed by a 5-point rolling mean filter. The 7-point median filter was used for b_{bp} profiles to account for large spikes in individual profiles caused by aggregations of particles (Briggs et al. 2011).

[Chl*a*] and b_{bp} profiles were then grouped into time series based on year of measurement. To account for seasonal polarity of North and South Hemispheres, dates for floats in the Southern Hemisphere were shifted forward six months prior to grouping. Floats generally complete vertical transits once every 5 to 10 days. Therefore, resultant time series were only included in analyses if the first profile was completed before January 20, the last profile after December 10, and if there were no more than 20 days between any two individual profiles.

To eliminate missing data in the final singular value decomposition (SVD) matrix, profiles in each time series were interpolated vertically to 1-meter intervals between 1 and 750 m. Each depth within the time series was then interpolated horizontally to 1-day intervals between 1 and 365 days. Interpolated [Chl*a*] and b_{bp} time series were then converted into vectors, creating 1 x 273,750 vectors for each parameter. [Chl*a*] and b_{bp} vectors were then

separately centered to mean 0, normalized to the standard deviation for all measurements in the vector, and joined end-on-end to create a 1 x 547,500 vector for each time series. This process was repeated for each complete time series in the data set, yielding a matrix consisting of 422 columns (corresponding to each time series in the data set) and 547,500 rows (corresponding to each day/depth in the time series). Including only depths up to 750 meters in the SVD matrix was found to maximize the proportion of overall variability captured in the first six singular vectors in SVD results. Note, however, that this choice did not preclude evaluating measurements from greater depths during subsequent analyses.

2.3.2. Cluster analysis to characterize similarities between time series

EOF was performed on the matrix of [Chl*a*] and b_{bp} time series, yielding a matrix of singular vectors capturing varying fractions of overall variability in the original time series matrix (Figure 2.1), and a matrix of coefficients reflecting the importance of each singular vector in each time series (Figure 2.2; see Appendix A.1 for details). To identify groups of similar time series, the pairwise Euclidean distance was calculated between subsets of the first six SVD coefficients (Pielou 1984). Similar groups of coefficients were then identified using divisive hierarchical clustering analysis—a clustering approach that has been indicated to provide more accurate clustering results than agglomerative methods, especially when a relatively small number of clusters is desired (Hastie et al. 2009).

In performing this analysis, several low-latitude time series, primarily in the Arabian Sea, were routinely grouped with high-latitude time series (Figure A.1). Therefore, to prevent misclassification without removing data, two rounds of clustering were performed. In the first round, time series were clustered into two groups based the first two coefficients in each column of U. Any time series with an absolute median latitude < 25° that was clustered with the high-latitude time series was then manually reassigned to the low-latitude cluster. This approach

effectively forced the segregation of high-latitude and low-latitude floats at the first clustering step, albeit without requiring any assumption of the appropriate boundary between the low-latitude and high-latitude floats.

Divisive hierarchical cluster analysis was then performed separately for time series in high-latitude and low-latitude clusters. Intracluster similarity was evaluated based on the average silhouette score for each cluster—a metric corresponding to the average distance between the coefficients for a given time series and the coefficients for all other time series within its assigned cluster (Rousseeuw, 1987; see S2 for details). The similarity of high-latitude time series was evaluated based on the Euclidean distance of coefficients corresponding to the first 2 singular vectors. The similarity of low-latitude time series was evaluated based on the Euclidean distance of coefficients corresponding to the first 6 singular vectors. Following this procedure, maximal silhouette scores were obtained for the high-latitude time series when two clusters were formed, while maximal silhouette scores were obtained for low-latitude time series when four clusters were formed.

2.3.3. Determination of environmental parameters

Environmental parameters derived for each profile are described in Table 2.1. Season was determined based on the dates after applying the adjustment for southern hemisphere floats. Winter, spring, summer and fall seasons, respectively, correspond to December – February, March – May, June – August, and September – November. For time series from regions subjected to seasonal monsoon forcing (see below), the year was divided into four periods: Northeast Monsoon (December 1 to April 1); Spring Intermonsoon (April 1 to July 1); Southwest Monsoon (July 1 to September 15); and Fall Intermonsoon (September 15 to November 1). Intermonsoon date ranges were centered on seasonal maxima in surface temperature, while monsoon date ranges were centered on seasonal minima in surface temperature (Figure A.2).

Mixed layer depth (MLD) was calculated as the depth where potential density increased above 0.03 kg m^{-3} in comparison to a reference density at 5 meters (de Boyer Montégut et al. 2004). The CANYON-B algorithm (Sauzède et al. 2017; Bittig et al. 2018) was used to obtain estimates of nitrate, phosphate, and silicate concentrations for floats equipped with dissolved oxygen sensors. For floats deployed in the Mediterranean sea, the CANYON-MED algorithm was used (Fourrier et al. 2020, 2021).

Estimated nutrient concentrations were used to calculate the depth and slope for the nitracline, phosphocline, and silicacline. The nitracline depth was defined as the depth where nitrate concentrations increased above a threshold of 1 μM relative to surface concentrations, following Lavigne et al. (2015). Nitracline slope was calculated by evaluating the rate of change in nutrients with depth between the nitracline depth and 1.25 times the nitracline depth (Cornec et al. 2021). The nitracline slope was interpreted as reflecting the upward vertical flux of nitrate at the depth of the nitracline (Gong et al. 2017). Parameters derived from estimated nutrient concentrations were strongly correlated with those derived from nutrient concentrations measured by float nitrate sensors (Pearson's r = 0.9, 0.7, and 0.87 for nitracline depth, nitracline PAR, and nitracline slope respectively; $p < 1 \times 10^{-16}$ for all comparisons). Appropriate thresholds for determining silicacline and phosphocline depths were determined for each profile as the ratio of maximal silicate or phosphate concentrations relative to the maximal nitrate concentrations. For example, if maximal silicate concentrations were 5 times greater than maximal nitrate concentrations, the silicacline would be defined as the depth where silicate concentrations increased above a threshold of 5 μM relative to surface concentrations.

The underwater light field for each profile was determined using float-based measurements of PAR. For floats not equipped with radiometers, PAR profiles were estimated

by estimating surface PAR based on the date, latitude, and longitude of a given profile using the clear sky model of Gregg & Carder (1990) and then propagating surface PAR estimates using attenuation coefficients estimated from the corresponding [Chl*a*] profiles following Morel et al. (2007). Estimated PAR profiles were then corrected to account for cloud cover, as described in Cornec et al. (2021). Daily integrated PAR was additionally estimated at each of the nutricline depths, following Lacour et al. (2017). Median daily integrated PAR within the mixed layer ($iPAR_{MLD}$) was derived for each profile by calculating daily integrated PAR at the median depth between the surface and MLD. The depth of the euphotic layer (Z_{eu}) was defined as the depth where PAR was reduced to 0.1% of its value at 1 meter. Following Dall'Ormo and Mork (2014), the depth of the productive layer (Z_p) was defined as the greater value of the MLD or Z_{eu} for a given profile.

Average [Chl*a*] and b_{bp} within the productive layer ($[Chla]_{pz}$ and bbp_{pz}, respectively) were calculated by integrating profiles from 1 m to the bottom of the productive layer and then normalizing to Z_p. Average b_{bp} values within the mesopelagic zone (bbp_{mz}) were calculated by integrating from Z_p to the bottom of the float transit (typically around 1000 m). Average b_{bp} values were also calculated for each of four equally sized depth ranges within the mesopelagic zone (bbp_{m1} – bbp_{m4}), defined for each profile as the difference between the bottom of the productive layer and the maximal depth of a given profile, divided by 4. To evaluate covariation between [Chl*a*] and b_{bp}, Spearman's correlation coefficient (ρ) was calculated for comparisons of $[Chla]_{pz}$ and bbp_{pz} for each biome. For ease of reporting, all b_{bp} values are reported in units of 10^{-3} m^{-1}.

2.3.4. Statistics

All analyses were performed in R 4.0.2 using RStudio 1.3.1073 (R Core Team 2016). Shapiro-Wilk tests were performed for each parameter to evaluate normality of data. Bartlett's tests were

used to evaluate homogeneity of variance. Because distributions of parameter measurements were generally not normally distributed, the median is used as the measure of central tendency for all results. The interquartile range (IQR) is used as the measure of dispersion, calculated as the difference between the third and first quartiles. instances where data were not normally distributed, Kruskal-Wallis tests were used to evaluate significant differences in parameter measurements based on cluster, season, or month of sampling, and Dunn's post-hoc test was used to perform multiple comparisons. For parameters where measurements were normally distributed, analysis of variance (ANOVA) tests were used, followed by Tukey's Honestly Significant Difference (HSD) post-hoc test. To evaluate main sources of variability within the dataset of environmental parameters, principal component analysis was performed on log-transformed and standardized parameter values. To evaluate relationships between pairs of parameters, linear regressions were performed on monthly parameter averages.

2.4. Results

2.4.1. Results of factor analysis

The first six singular vectors identified in SVD analysis captured 81 percent of the variability in the dataset of 422 time series (Figure 2.1, Table A.2). The contribution of each vector to the original time series depends on both the magnitude and sign of the corresponding coefficient (Figure 2.2). For example, the first singular vector alone, when multiplied by a negative coefficient, resulted in a time series characterized by large surface [Chl*a*] and b_{bp} values during summer months. The second singular vector, when multiplied by a negative coefficient, counteracted the effect of the first singular vector and resulted in time series characterized by a seasonal deep chlorophyll maximum (DCM). If multiplied by a positive coefficient, the second singular vector would enhance the contribution of the first singular vector. In this way, each

coefficient adds additional information about the vertical and seasonal distribution of [Chl*a*] and b_{bp} in the original time series and can be used to differentiate time series via cluster analysis.

2.4.2. Results of cluster analysis

The initial clustering (Figure 2.3A) formed a low-latitude cluster with a unimodal latitudinal distribution centered on 7.4°N ± 24.8°N (n = 190) and a high-latitude cluster with a bimodal latitudinal distribution centered on 58.0°N ± 8.2°N and -50.24°N ± 6.85°N (n = 232). High-latitude time series were further divided (Figure 2.3B, Figure A.3) into a high-chlorophyll bloom (HCB) cluster centered at 60.4°N ± 7.2°N (n = 164) and a low-chlorophyll bloom (LCB) cluster centered at -49.9°N ± -6.7°N (n = 68). Low-latitude time series were divided into Arabian Sea (AR), Equatorial (EQ), Permanent DCM (PDCM) and Seasonal DCM (SDCM) clusters. AR time series were centered at 16.01° ± 2.57° N (n = 18); EQ time series were centered at 4.18° N ± 13.08° (n = 60); PDCM time series had a bimodal distribution centered at 23.68° ± 3.35° N and -22.06° ± 4.05° N (n = 49); SDCM time series had a bimodal distribution centered at 36.01° ± 3.20° N and -34.98° ± 8.10° N (n = 63). SDCM time series in northern latitudes were nearly entirely limited to the Mediterranean Sea, while those in southern latitudes were concentrated at latitudes separating the southern subtropical gyres from the Southern Ocean. To capture any differences in environmental drivers between these regions, Mediterranean (hereafter MED) and open ocean SDCM time series (hereafter Subtropical Frontal Zone, or STFZ) were analyzed separately in addition to those corresponding to the SDCM time series as a whole. Following Longhurst (2007), in the remainder of the text the HCB, LCB, SDCM, PDCM, AR, and EQ clusters are referred to as biomes, since each has a characteristic seasonal phytoplankton cycle (see below). High-latitude clusters and low-latitude clusters are still identified as clusters since they are themselves composed of several biomes.

2.4.3. Trends in bio-optical parameters for each biome

Averaging the time series in matrix M corresponding to each biome revealed distinct seasonal trends in [Chl*a*] (Figure 2.4, Table 2.2) and b_{bp} (Figure 2.5, Table 2.3). Time series in the EQ biome were characterized by a DCM at a median depth of 60.5 m (IQR = 20.5 m) that persisted throughout the year. Time series in the AR biome were characterized by the formation of a DCM during intermonsoon periods at median depths of 55.5 m (IQR = 18.4 m) and 34.5 m (IQR = 15.0 m) during the Spring Intermonsoon and Fall Intermonsoon, respectively. Time series in the SDCM biome were characterized by a spring phytoplankton bloom, followed by the formation of a DCM that reached a maximal depth of 91.0 m (IQR = 31.5 m) in summer months. Time series in the PDCM biome were characterized by a permanent DCM that reached a maximal depth of 136.0 m (IQR = 28.0 m) during summer months. Time series in the LCB and HCB biomes were both characterized by the formation of summer phytoplankton blooms, without the formation of seasonal DCMs. Annually averaged $[\text{Chl}a]_{sat}$ values were highest for HCB (median = 0.48, IQR = 0.54), AR (median = 0.32, IQR = 0.27), and LCB biomes (median = 0.24, IQR = 0.18). Values were considerably lower and less variable for the EQ (median = 0.13, IQR = 0.06) and SDCM (median = 0.12, IQR = 0.11) biomes. PDCM time series represented global minima (median = 0.05, IQR = 0.03).

Trends in b_{bp} generally mirrored those in [Chl*a*] in HCB, LCB, and AR biomes, although showed comparatively less variability with depth in EQ, SDCM, and PDCM biomes. Spearman's correlation coefficients for comparisons of bbp_{pz} and $[\text{Chl}a]_{pz}$ were 0.81 for the HCB biome, 0.78 for the LCB biome, and 0.77 for the AR biome. Relationships between bbp_{pz} and $[\text{Chl}a]_{pz}$ were more variable in SDCM, EQ, and PDCM biomes, with correlation coefficients of 0.67, 0.69, and 0.21, respectively. Within the SDCM biome, correlation coefficients were 0.83 for STFZ time

series and 0.58 for MED time series. Correlation coefficients were statistically significant. with p-values of less than 1 x 10^{-24} for all biomes.

While maximal b_{bp} values were located within the productive zone, there were frequently elevated b_{bp} values within the mesopelagic zone as well, especially during summer months. In the AR biome in particular, b_{bp} values reached local minima near the base of the productive zone (see below) and increased to local maxima with further increases in depth (Figure 2.5, Figure A.4). There were additionally pronounced inter-biome differences in depth-integrated b_{bp} values in the upper and lower mesopelagic. In the EQ and AR biomes, bbp_{m1} values showed limited seasonal variability, with annual median values of 0.33 m^{-1} (IQR = 0.10 m^{-1}) in EQ and 0.51 m^{-1} (IQR = 0.18 m^{-1}) in AR. In the SDCM biome bbp_{m1} values reached maxima during of 0.37 m^{-1} (IQR = 0.13 m^{-1}) during the spring bloom. The LCB and HCB biomes showed pronounced seasonality in bbp_{m1} values, increasing to 0.50 m^{-1} (IQR = 0.20 m^{-1}) and 0.55 m^{-1} (IQR = 0.28 m , respectively, during the summer bloom. Differences in bpp_{m4} values between biomes were much less pronounced, with values in the EQ, SDCM, LCB, and HCB biomes being extremely similar (overall median = 0.25 m^{-1}, IQR = 0.09 m^{-1}). bbp_{m4} values were comparatively elevated in the AR biome (median = 0.33 m^{-1}, IQR = 0.05 m^{-1}) and reduced in the PDCM biome (median = 0.55 m^{-1}, IQR = 0.06 m^{-1}).

2.4.4. Global variability in environmental parameters

Results of principal component analysis performed on environmental parameters captured 1) the degree of correlation between individual environmental parameters and 2) main sources of variability in environmental parameters for each biome (Figure 2.6). The first two principal components represented 75.8 % of the total variability in the data set. Profiles in HCB and LCB biomes were associated with high nitrate concentrations within the mixed layer and low light levels at the depth of the nitracline. Overall, profiles in AR and EQ biomes were characterized by

low nitrate concentrations within the mixed layer, elevated light levels at the depth of the nitracline, and steep nitracline slopes. Profiles in SDCM and PDCM biomes were characterized by deep nitraclines, low light levels at the depth of the nitracline, shallow nitracline slopes, and high surface salinities.

2.4.5. Environmental setting of the Equatorial biome:

Time series in the EQ biome were characterized by limited seasonal variability in environmental parameters (Figure A.5). Time series had a median surface temperature of 28.7 °C (IQR = 2.2 °C), and a median MLD of 25.0 m (IQR = 26.0 m). The nitracline was situated 29.0 m (IQR = 32 m) beneath the MLD throughout the year, and median nitrate concentrations within the mixed layer were perennially low (median = 0.2 μM, IQR = 0.9 μM). Values for nitracline slope in EQ time series represented global maxima, with an annual median value of 0.62 μM m^{-1} (IQR = 0.69 μM m^{-1}). $iPAR_{NO3}$ were high throughout the year (median = 0.91 E m^{-2} d^{-1}, IQR = 1.62 E m^{-2} d^{-1}).

2.4.6. Environmental setting of the Arabian Sea biome

Time series in the AR biome were characterized by bimodal seasonality in virtually all parameters (Figure A.6). Surface temperatures ranged from median values of 26.3 °C (IQR = 1.2 °C) during the Northeast Monsoon, to median values of 29.6 °C (IQR = 1.7 °C) during Summer Intermonsoon. Mixed layer depths reached minimal depths of 11.0 m (IQR = 6.00 m) and 12.0 m (IQR = 9.0 m) during the Spring Intermonsoon and Fall Intermonsoon, respectively, and increased to 48.0 m (IQR = 41.0 m) and 45.0 m (IQR = 25.0 m) during the Northeast and Southwest Monsoon, respectively.

The nitracline was situated 26.0 m (IQR = 17.0 m) beneath the MLD during the Spring Intermonsoon and 14.0 m (IQR = 17.0 m) beneath the MLD during the Fall Intermonsoon. In contrast, the MLD and nitracline depths became coincident during the Northeast Monsoon and

Southwest Monsoon. Nitrate concentrations in the mixed layer increased during monsoon periods, reaching annual maxima of 3.1 μM (IQR = 3.0 μM) and 2.4 μM (IQR = 3.1 μM) during the Northeast Monsoon and Southwest Monsoon, respectively. Nitrate concentrations in the mixed layer decreased to concentrations of 1.2 μM (IQR = 1.8 μM) and 1.7 μM (IQR = 3.0) during Spring Intermonsoon and Fall Intermonsoon, respectively.

The nitracline slope was generally high throughout the year (median = 0.38 μM m^{-1}, IQR = 0.46 μM m^{-1}), although decreased sharply at the onset of Spring Intermonsoon period before increasing to annually maximal values during the Southwest Monsoon period. Values for $iPAR_{NO3}$ during intermonsoon periods represented global maxima, reaching 2.2 E m^{-2} d^{-1} (IQR = 3.2 E m^{-2} d^{-1}) during the Spring Intermonsoon and 1.1 E m^{-2} d^{-1} (IQR = 2.0 E m^{-2} d^{-1}) during the Fall Intermonsoon. Values for $iPAR_{NO3}$ were comparatively low during monsoon periods, decreasing to 0.1 E m^{-2} d^{-1} (IQR = 0.4 E m^{-2} d^{-1}) during the Southwest Monsoon and to 0.14 E m^{-2} d^{-1} (IQR = 1.0 E m^{-2} d^{-1}) during Northeast Monsoon.

2.4.7. Environmental setting of the Seasonal DCM biome

Time series in the SDCM biome were characterized by strong seasonality in all environmental parameters (Figure A.7). Surface temperatures in SDCM time series ranged from median values of 16.7 °C during winter months to 23.7 °C during summer months. The MLD reached a maximal depth of 72.0 m (IQR = 66.5 m) in January and shoaled to 13.0 m (IQR = 15.0 m) in summer months. The nitracline was located 38.0 m (IQR = 69.0 m) beneath the MLD during winter months, and 83.0 m (IQR = 64.0 m) during summer months. The nitracline slope was low compared to other biomes and showed limited seasonal variability (median = 0.05, IQR = 0.07). Nitrate concentrations within the mixed layer showed a small degree of seasonal variability, reaching maximal concentrations of 0.27 μM (IQR = 0.73 μM) during winter months, and decreasing to 0.07 μM (IQR = 0.20 μM) during summer months. Median values of $iPAR_{NO3}$

showed limited seasonal variability, ranging from 0.008 E m^{-2} d^{-1} (IQR = 0.03 E m^{-2} d^{-1}) during winter months to 0.19 E m^{-2} d^{-1} (IQR = 0.56 E m^{-2} d^{-1}) during summer months.

Separately analyzing MED (Figure A.8) and STFZ (Figure A.9) time series revealed distinct hydrological and biogeochemical dynamics within the SDCM biome. The extent of winter mixing was significantly greater (ANOVA, $p < 2 \times 10^{-16}$) in STFZ time series (median = 106.0 m, IQR = 51.0 m) compared to MED time series (median 61.0 m, IQR = 57.0 m). In STFZ time series, nitrate concentrations within the mixed layer ranged from maxima of 0.70 µM (IQR = 1.12 µM) during winter months to 0.06 µM (IQR = 0.23 µM). In MED time series N_{MLD} values were comparatively reduced, ranging from 0.19 µM (IQR = 0.34 µM) during winter months to 0.07 µM (IQR = 0.15 µM) during summer months. Differences in estimated phosphate concentrations within the mixed layer were more pronounced, reaching maxima of 0.24 µM (IQR = 0.14 µM) during winter months in STFZ time series, compared to maxima of 0.02 µM (IQR = 0.01 µM) during winter months in MED time series.

2.4.8. Environmental setting of the Permanent DCM biome

Time series in the PDCM biome were characterized by intermediate seasonality in all environmental parameters and a permanent separation of the nitracline and MLD (Figure 2.4). Surface temperatures ranged from median values of 24.3 °C (IQR = 2.3 °C) in spring months to 27.3 °C (IQR = 1.6 °C) in fall months. Mixed layer depths ranged from 21.0 m (IQR = 20.0 m) during summer months to 78.5 m (IQR = 51.0 m) in winter months. The nitracline was located 69.0 m (IQR = 56.25 m) beneath the MLD throughout the year, ranging from 69.0 m (IQR = 56.25 m) in winter months to 141.0 m (IQR = 52.0 m) in summer months. There were no PDCM profiles where the MLD was situated below the nitracline depth. Nitrate concentrations were depleted in the mixed layer throughout the year, representing global minima (median = 0.1 µM,

IQR = 0.14). Median values for $iPAR_{NO3}$ ranged from 0.014 E m^{-2} d^{-1} (IQR = 0.04 E m^{-2} d^{-1}) during fall months to 0.038 E m^{-2} d^{-1} (IQR = 0.08 E m^{-2} d^{-1}) during spring months.

2.4.9. Environmental setting of the Low Chlorophyll Bloom biome

Time series in the LCB biome were characterized by deep winter mixing and comparatively weak summer stratification (Figure A.11). Surface temperatures showed comparatively little seasonal variability, ranging from 5.8 °C (IQR = 6.6 °C) during winter months to 8.1 °C during summer months (IQR = 8.15 °C). The MLD increased to 146 m (IQR = 87.5 m) during winter months and shoaled to 52 m (IQR = 47.0 m) during summer months. The nitracline depth was roughly coincident with the MLD throughout the year, located 6.9 m beneath the MLD on average (IQR = 7.4 m). Average nitrate concentrations showed limited annual variability, reaching maximal values of 21.6 µM (IQR = 13.3 µM) during spring months, and decreasing to 19.3 µM (IQR = 15.6 µM) during summer months. iPARMLD values increased from 0.04 E m-2 d-1 (IQR = 0.14 E m-2 d-1) during winter months to 2.8 E m-2 d-1 (IQR = 6.9 E m-2 d-1) during summer months.

2.4.10. Environmental setting of the High Chlorophyll Bloom biome

Time series in the HCB biome were characterized by deep winter mixing and a degree of summer stratification similar to that in low-latitude floats (Figure A.12). Surface temperatures represented global minima, ranging from 4.9 °C (IQR = 3.3 °C) during winter months to 8.6 °C (IQR = 4.4 °C) during summer months. Annual variability in MLD was greater than in any other biome, reaching maximal depths of 165 m (IQR = 139 m) during winter months and shoaling to 23 m (IQR = 23 m) during summer months. As in the LCB biome, the nitracline depth was generally coincident with the MLD throughout the year. Nitrate concentrations within the mixed layer were more variable than those in LCB time series, decreasing from 15.9 µM (IQR = 15.2 µM) during winter months to 5.4 µM (IQR = 18.9 µM) during summer months. $iPAR_{MLD}$ values

increased from 0.015 E m^{-2} d^{-1} (IQR = 0.05 E m^{-2} d-1) during winter months to 5.1 E m^{-2} d^{-1} (IQR = 7.5 E m^{-2} d^{-1}) during summer months.

2.5. Discussion

2.5.1. Distribution of biomes and comparison with existing classifications

We identified six biomes within the global ocean based on time series of bio-optical measurements from BGC-Argo floats. Results of cluster analyses indicate that this grouping provides the greatest possible intracluster similarity in EOF weightings and, by extension, in the original [Chla] and bbp time series. Clustering EOF weightings into smaller or larger numbers of groups reduced silhouette scores in either case. Dividing time series into a smaller number of clusters also resulted in less geographically or ecologically interpretable results, for example leading to the fusion of EQ and PDCM biomes when grouping low-latitude time series into 3 clusters rather than 4. Conversely, grouping low-latitude floats into 5 time series divided the SDCM biome into two clusters differentiated by the depth of the DCM during summer months, with the new clusters having lower silhouette scores than the original SDCM biome. Although our results required the manual reclassification of Arabian Sea time series and two Equatorial time series, doing so ultimately increased the silhouette scores of the final clusters obtained, indicating that the manual reassignment leads to greater intracluster similarity than would otherwise be possible.

The clusters obtained in our analysis roughly correspond to five of the eight ecological models described by Longhurst (1995). Briefly, these models are 1) Polar irradiance mediated-production peak, 2) Nutrient limited spring production peak, 3) Winter/spring production with nutrient limitation, 4) Small amplitude response to trade wind seasonality, and 5) Large amplitude response to monsoon reversal. The other three models correspond to coastal environments excluded from our analysis. The HCB and LCB biomes differentiate between

differing phytoplankton dynamics within model 1. The SDCM and PDCM biomes correspond to models 2 and 3, respectively. The EQ biome generally corresponds to model 4, albeit with some exceptions corresponding to the formation of shallow summer DCMs in several Southern Ocean time series. The AR biome, limited entirely to the Arabian Sea, corresponds to model 5.

Uniquely, however, in our analysis we were able to objectively identify the distribution these biomes based on subsurface phytoplankton bio-optical measurements alone. Because spatiotemporal variability in these measurements integrate the top-down (e.g., grazing) and bottom-up (e.g., resource limitation) processes influencing phytoplankton distributions within each biome, the coefficients derived from EOF analysis provide informative criteria than would subsets of physical, hydrological, or biogeochemical parameters. Indeed, our PCA results highlight the difficulty in recreating the obtained biomes based on environmental parameters alone: while the biomes can be differentiated into pairs coarsely reflecting their latitudinal distribution (e.g. HCB and LCB, SDCM and PDCM, EQ and AR), there was insufficient structure in the environmental data to further differentiate between profiles belonging to individual biomes (Figure 2.6).

2.5.2. Environmental drivers of phytoplankton distributions in high-latitude biomes

The high-latitude HCB and LCB biomes were both characterized by summer phytoplankton blooms that coincided with 1) an increase in light levels at the depth of the nitracline, and 2) a gradual drawdown of nitrate within the mixed layer, especially in HCB. The main factor differentiating time series within high-latitude biomes was the magnitude of the seasonal contrast in vertical [Chl*a*] distributions. $[\text{Chl}a]_{pz}$ values in the HCB time series underwent a 3-fold increase between winter and summer months, compared to less than a 2-fold increase in LCB time series (Figure 2.4 Table 2.2). Similarly, $[\text{Chl}a]_{sat}$ values in the HCB biome underwent a 3-fold increase between winter and summer months, while those in the LCB biome were

comparatively constant throughout the year. Given that the majority of the LCB time series are located within the Southern Ocean, the reduced magnitude of summer $[Chl\mathit{a}]_{pz}$ and $[Chl\mathit{a}]_{sat}$ values in the LCB biome likely reflects limited iron availability within the productive zone (Boyd and Doney 2002; Venables and Moore 2010; Moore et al. 2013). The differences between annual variability in $[Chl\mathit{a}]_{pz}$ compared to $[Chl\mathit{a}]_{sat}$ within the LCB biome could be attributed in part to increased fluorescence per unit Chla under iron limitation (Behrenfeld et al. 2006b, 2009), which could increase float measurements of [Chl*a*] (based on chlorophyll fluorescence) compared to those made via satellite (based on ocean color).

Differences in [Chl*a*] distributions in the HCB and LCB biomes could also be due to seasonal differences in the extent of vertical mixing. Mixed layer depths were greater in HCB time series during winter months and greater in LCB time series during summer months (Figure 2.4). Deeper winter mixing in the HCB time series would result in lower temperatures and longer periods of darkness for phytoplankton in HCB time series, both of which could reduce growth rates (Walter et al. 2015) and the concentration of algal cells in the mixed layer (Evans et al., 1985). The same mechanism could reduce [Chl*a*] in the LCB time series during summer months. Finally, the shallower mixed layer depths associated with HCB time series during summer months could contribute to the elevated growth irradiances observed within the mixed layer during periods of maximal stratification (Sverdrup 1953; Obata et al. 1996).

Despite the HCB and LCB biomes being clearly differentiated by the clustering algorithm (primarily due to the much larger weightings associated with the second singular vector in HCB time series compared to LCB time series (Figure 2.2), there was significant overlap of LCB and HCB time series in the Southern Ocean (Figure 2.3, Figure A.3). The occurrence of time series belonging to both biomes within in the Southern Ocean is likely due to interannual or regional

variability in bloom intensity, perhaps in response to localized increases in iron concentrations along the sampling track of individual floats, hydrothermal vent activity (Ardyna et al. 2019; Schine et al. 2021), atmospheric deposition (Erickson et al. 2003), advection of iron-enriched water (Ellwood et al. 2008; Sedwick et al. 2008), or inputs from islands (Blain et al. 2007). Given that HCB time series in the Southern Ocean were somewhat concentrated within higher latitudes, their occurrence may also reflect local mitigation in nutrient limitation due to continental runoff or upwelling of bottom water (de Baar et al. 1995; Klunder et al. 2011).

2.5.3. Environmental drivers of phytoplankton distributions in low-latitude biomes

In contrast to the summer bloom of high-latitude biomes, low-latitude biomes were broadly characterized by low nutrient concentrations within the mixed layer (Figure 2.6) and the formation of a DCM at the depth of the nitracline following the onset of stratification. With the exception of the AR biome, the main factor differentiating low-latitude biomes was the depth of the DCM (Figure 2.4), as reflected by the magnitude and sign of weightings for the third singular vector in the EQ (large positive values), SDCM (values near zero), and PDCM (negative values) time series (Figure 1, Figure 2).

2.5.4. Environmental drivers of phytoplankton distributions in Seasonal DCM and Permanent DCM biomes

Beyond maximal DCM depth, SDCM time series were additionally differentiated from PDCM time series by the occurrence of a prominent spring bloom, as reflected by seasonal increases in $[\text{Chl}a]_{\text{pz}}$ in SDCM time series (Figure 2.4, Table 2.2). It is worth noting, though, that despite similar seasonal trends in bio-optical parameters, the SDCM biome represents an admixture of time series from two geographically distinct systems. The STFZ time series captures a transition zone between the nutrient-limited waters of the central gyres and the nutrient-rich waters of the Southern Ocean, as reflected in part by globally intermediate nutrient concentrations within the mixed layer during winter months (Smythe-Wright et al. 1998; Spalding et al. 2012). In contrast,

time series in the Mediterranean capture comparatively nutrient limited conditions, especially with respect to PO_4, reflecting the near-oligotrophic status of the Mediterranean (D'Ortenzio and Ribera d'Alcalà 2008), especially in the Eastern Basin. There were also a greater proportion of profiles where the MLD exceeded the nitracline depth during winter mixing in STFZ time series compared to MED time series (data not shown), highlighting the importance of advective nutrient transport to spring bloom dynamics in the Mediterranean (Lavigne et al. 2013; D'Ortenzio et al. 2014).

As a result of reduced nutrient availability, it is also likely that seasonal [Chl*a*] variability in MED time series is more strongly influenced by photoacclimation processes, whereby phytoplanktonic cells increase cellular concentrations of photopigments in response to light limitation (Fennel and Boss 2003; Letelier et al. 2004; Mignot et al. 2014a). In our results, relationships between $[Chl\mathit{a}]_{pz}$ and bbp_{pz} were considerably weaker in MED time series compared to STFZ time series. This is consistent with previous studies, which have demonstrated photoacclimation to play a major role in seasonal chlorophyll dynamics in the Mediterranean, especially in the Eastern basin, both during the spring bloom (Bellacicco et al. 2016) and in the formation of the DCM (Cornec et al. 2021).

In contrast to the SDCM time series, there were only minimal winter increases in $[Chl\mathit{a}]_{pz}$ in PDCM time series, despite maximal mixed layer depths in PDCM time series being similar or greater than those in SDCM time series. This underlines that phytoplankton biomass in oligotrophic regions (e.g., areas characterized by low levels of surface [Chl*a*] and dominated by the subtropical gyres) is not limited by weak vertical mixing so much as by a perennially deep nitracline inaccessible to mixing compared to other biomes (McClain et al. 2004). Globally minimal $iPAR_{NO3}$ values at the depth of the nitracline in the PDCM biome, along with weak

correlations between bbp_{pz} and $[Chl\mathit{a}]_{pz}$, again suggest the critical importance of photoacclimation processes at the depth of the DCM and above in subtropical gyres as reported in previously (Mignot et al. 2014b; Cornec et al. 2021). The homogenous b_{bp} distributions in PDCM time series compared to unimodal distributions of [Chl*a*] (Figure 2.4, Figure 2.5) likely also reflect the dominance of the b_{bp} signal by non-algal biomass within the euphotic zone, as has been previously documented in oligotrophic regions for both attenuation coefficients (Claustre et al. 1999; Grob et al. 2007b) and b_{bp} (Bellacicco et al. 2019).

2.5.5. Environmental drivers of phytoplankton distributions in EQ and AR biomes

The primary factor distinguishing phytoplankton distributions in EQ and AR biomes was unambiguously the influence of the monsoon system in the AR biome, which results in the marked bimodal seasonality in vertical distributions of [Chl*a*] and b_{bp} in AR time series and is reflected by the positive values for singular vectors 4 and 6 in AR time series (Figure 2.1, Figure 2.2). During monsoon periods, the observed increases in $[Chl\mathit{a}]_{pz}$ and bbp_{pz} are generally regarded to be attributable to upwelling, resulting in elevated nitrate concentrations within the mixed layer and the coincidence of the MLD and nitracline (Bauer et al. 1991). Conversely, shoaling of the MLD during the intermonsoon results in a rapid depletion of nitrate near the surface, accompanied by the formation of a DCM (Mccreary et al. 1996; Ravichandran et al. 2012; Prasanth et al. 2021). By contrast, [Chl*a*] and b_{bp} distributions in the EQ biome are representative of typical tropical structure, characterized by limited seasonal variability in all environmental parameters, resulting in a persistent DCM coincident with the depth of the nitracline (Herbland and Voituriez 1979). It is worth highlighting that time series in the Bay of Bengal were grouped with EQ time series instead of AR time series. This reflects the weakened influence of monsoon cycling on phytoplankton distributions in the Bay of Bengal compared to the Arabian Sea, likely due to elevated stratification and reduced upwelling throughout the year

in the Bay of Bengal (Gauns et al. 2005). Indeed, any monsoon influence for the Bay of Bengal floats is not readily discernable in the average [Chl*a*] or b_{bp} time series for the EQ biome (Figure 2.4, Figure 2.5).

Despite these differences, the DCM in the EQ and AR biome are both characterized by high light levels and steep NO_3 gradients at the depth of the nitracline. This is in contrast to the comparatively low-light and weak nutrient flux at the DCM in SDCM and PDCM biomes (Figure 2.6), and likely corresponds to a greater degree of biomass accumulation at depth than would be otherwise possible (Herbland and Voituriez 1979; Beckmann and Hense 2007; Cullen 2015). In the AR biome, these factors combined with additional pulses of nutrients via upwelling make the Arabian Sea one of the most productive regions in the global ocean.

2.5.6. Drivers of biomass accumulation in the productive zone

Within this general framework, evaluating relationships between bbp_{pz} and environmental parameters revealed specific drivers of phytoplankton dynamics in each biome. Specifically, strong relationships between $iPAR_{MLD}$ and bbp_{pz} in high latitude time series provide strong evidence of light limitation in HCB and LCB biomes, where changes in $iPAR_{MLD}$ coincided with the onset and decay of the summer bloom. Notably, however, the LCB biome showed more gradual initial increases in biomass accumulation with increases in $iPAR_{MLD}$, in addition to lower maximal bbp_{pz} values. Provided that seasonal increases in b_{bp} primarily correspond to increases in algal biomass in high latitude biomes (as is suggested by strong correlations between bbp_{pz} and $[Chl\textit{a}]_{pz}$; also see (Bellacicco et al. 2019; Rembauville et al. 2017) this seems to capture nutrient limitation at the basin scale. In particular, the reduced initial slope observed in the LCB biome reflects experimental and field studies demonstrating iron limitation to reduce photosynthetic efficiency at low light intensities (Boyd 2002; Strzepek, Boyd, and Sunda 2019; Sunda and Huntaman 1997). At the peak of the bloom, phytoplankton growth may be co-limited

by both light and nutrients, given that maximal monthly averaged $iPAR_{MLD}$ values in the LCB biome are only half those in the HCB biome (Figure 2.7, Figure A.11, Figure A.12).

The contrasting relationships between $iPAR_{NO3}$ and bbp_{pz} at low-latitudes reflects the range of seasonal dynamics characterizing the EQ, AR, PDCM, and SDCM biomes (Table A.4). The slightly negative, albeit statistically significant relationship between $iPAR_{NO3}$ and bbp_{pz} in the AR biome likely reflects primarily nutrient limited conditions within the Arabian Sea and the importance of sustained nutrient delivery via upwelling to biomass accumulation during monsoon periods (Cullen et al. 2002; Chowdhury et al. 2021). This contrasts with high-latitude blooms, where seasonal detrainment with the shoaling of the mixed layer results in positive relationships between $iPAR_{NO3}$ and bbp_{pz}. The exceptionally weak coupling of bbp_{pz} and nitracline parameters in the SDCM time series may result from several factors: 1) the separation of the MLD and nitracline during periods of maximal biomass accumulation (e.g. the spring bloom), resulting in phytoplankton populations supported by external nutrient inputs; 2) a large degree of seasonal variability in relationships between b_{bp} and non-algal particles, especially in the MED time series; 3) variable phytoplankton dynamics both within and between the MED and SFTM time series.

2.5.7. Biomass accumulation and attenuation in the mesopelagic zone

Despite these differences in the magnitude and seasonality in upper ocean biomass accumulation, global variability in biomass accumulation in the mesopelagic zone was half that of biomass accumulation in the productive zone (Figure 2.8A & 2.8B). This is primarily the result of distinct seasonal trends in biomass accumulation across biomes, in addition to inter-biome differences in attenuation of the sinking particulate material and hence of b_{bp} signal within the mesopelagic zone. First, while HCB and LCB biomes are characterized by marked seasonality in bbp_{mz} values and globally maximal bbp_{mz} values during the summer bloom, intermediate and seasonally

invariable bbp_{mz} values in EQ time series result in relatively similar annual bbp_{mz} distributions across the three biomes. Secondly, and particularly in high latitude time series, attenuation rates differ between HCB and LCB biomes, resulting in similar bbp_{m2} values despite different bbp_{m1} values during the summer bloom (Figure 2.8C). Beyond minimizing differences in bbp_{mz} between HCB and LCB biomes, this may additionally reflect enhanced export efficiency in LCB time series, potentially due to higher proportions of picophytoplankton and nanophytoplankton groups in the HNLC regions of the Southern Ocean (Bouman et al. 2012; Boyd 2002; Guidi et al. 2009).

Finally, the comparatively muted seasonality in biomass accumulation and attenuation in the AR biome is striking given the magnitude of the monsoon-induced variability in bio-optical parameters observed in the euphotic zone (Figure 2.4, Figure 2.5). This may be in part due to the elevated b_{bp} measurements at the depth of the oxygen minimum zone (Figure A.4; also visible as the local b_{bp} maximum centered on ~300 m in Figure 2.5), which are potentially attributable to populations of denitrifying bacteria (Roullier et al. 2014; Ward et al. 2009; Whitmire et al. 2009). If the formation and magnitude of the OMZ were due at least in part to processes decoupled from euphotic zone processes (e.g., weak advection within the mesopelagic zone; see Wiggert 2005 for a review of proposed mechanisms), these bacterial populations could obscure seasonality in biomass accumulation attributable to phytoplankton carbon. This seems especially likely given the comparatively enhanced seasonality in the lower mesopelagic zone (bbp_{m3} and bbp_{m4} on figure 2.8C), which is characterized by maxima and minima in biomass accumulation corresponding to the monsoon and intermonsoon periods, respectively.

2.6. Conclusion

To date, characterizing phytoplankton distributions at the global scale has been limited by the near-invisibility of subsurface phytoplankton distributions to remote sensing platforms. As the issue was described by Platt and Sathyendranath, "to characterize the ocean water column using satellite-based observations of its surface layer is certainly not ideal. But is the only option available at these scales, now and into the foreseeable future" (1988). However, with modern advances, our results demonstrate that the amount and extent of data collected by BGC-Argo profiling floats is now sufficient to resolve this constraint.

Beyond demonstrating the utility of data from BGC-Argo floats for partitioning biogeographical provinces within the global ocean, this analysis provides new insight into the drivers of subsurface variability in phytoplankton distributions at the global scale, while also highlighting the role of subsurface processes in driving biomass accumulation in the mesopelagic zone. Our results demonstrate the diversity of biogeochemical and hydrological interactions shaping phytoplankton distributions in the global ocean, especially highlighting 1) the importance of light level at the depth of the nitracline as driver of euphotic zone biomass accumulation in low and mid-latitudes, 2) patterns of co-limitation of light and nutrients in the onset of the summer bloom in the Southern ocean, and 3) globally variable rates of biomass attenuation and accumulation in the mesopelagic zone.

By capturing seasonal trends in subsurface phytoplankton distributions, our results additionally provide a novel "template" for organizing and interpreting existing oceanographic data sets. While we did not do so here, profile measurements of phytoplankton community structure and/or HPLC pigments from ship-based research programs (e.g. JGOFS, AMT, HOT) could be used to more fully inform subsurface phytoplankton dynamics within each of the

identified biomes. Similarly, our results could be useful for constraining estimates of global primary production by constraining estimated subsurface phytoplankton proxies in primary production models. Either of these goals would benefit from extrapolating the results of our analysis to regions not directly captured by the BGC-Argo network in its present configuration. Toward this end, the distinct spatial distribution of time series in each biome could be used to constrain the geographic distributions of biomes based on variance in surface measurements. For instance, we found a random forest classifier trained on statistical parameters derived from MODIS data for each profile, in addition to latitude and longitude, to correctly predict biome membership in a training data set in more than 96% of cases. And while our current data set is too sparse to apply this classifier to historical data with informative results, such a goal will surely become feasible in the near-future with the increasing deployment of additional BGC-Argo floats, especially in regions somewhat under sampled up to now.

Table 2.1: Abbreviations used in chapter 2

Symbol	Definition	Units
AR	Arabian Sea biome	
EQ	Equatorial biome	
HCB	High chlorophyll bloom biome	
LCB	Low chlorophyll bloom biome	
PDCM	Permanent deep chlorophyll maximum biome	
SDCM	Seasonal deep chlorophyll maximum biome	
[Chl*a*]	Chl*a* concentration	mg m^{-3}
$[Chl a]_{DCM}$	Chl*a* concentration at the deep chlorophyll maximum	mg m^{-3}
$[Chl a]_{mz}$	Average Chl*a* concentration within the mesopelagic zone	mg m^{-3}
$[Chla]_{zp}$	Average Chl*a* concentration within the productive zone	mg m^{-3}
$[Chl a]_{sat}$	Satellite [Chl*a*]	mg m^{-3}
b_{bp}	Particulate backscattering coefficient	m^{-1}
bbp_{zp}	Average b_{bp} within the productive zone	m^{-1}
bbp_{mz}	Average b_{bp} within the mesopelagic zone	m^{-1}
iPAR	Daily integrated PAR	E m^{-2} d^{-1}
$iPAR_{MLD}$	Median iPAR in the mixed layer	E m^{-2} d^{-1}
$iPAR_{NO3}$	iPAR at the nitracline depth	E m^{-2} d^{-1}
MLD	Mixed layer depth	m
$N2_{MLD}$	Brunt-Vaisala frequency at the mixed layer	HZ
N_{MLD}	Average NO_3 concentration within the mixed layer	µM
PAR	Photosynthetically available radiation	E m^{-2}
P_{MLD}	Average PO_4 concentration within the mixed layer	µM
S_{MLD}	Average SiO_4 concentration within the mixed layer	µM
Z_{DCM}	Depth of deep chlorophyll maximum	m
Z_{ez}	Depth of euphotic zone; depth where PAR is reduced to 0.1% of its surface value	m
Z_{NO3}	Depth of the nitracline	m
Z_{pz}	Depth of productive zone; maximum of MLD or z_{eu} for a given profile	m

Table 2.2: Average [Chl*a*] within the productive zone for each biome

Average Chl*a* concentrations (mg m^{-3}) within the productive zone for each biome identified in SVD analysis. All values are calculated by interpolating profiles to 1-meter intervals and normalizing to depth of the productive zone. Values are reported as the median (IQR). Note that in AR time seires, Winter, Spring, Summer and Fall correspond to Northeast Monsoon, Summer Intermonsoon, Southwest Monsoon, and Fall Intermonsoon, respectively.

	Depth integrated Chl*a*			
	Winter	**Spring**	**Summer**	**Fall**
EQ	0.20 (0.11)	0.21 (0.10)	0.20 (0.8)	0.23 (0.8)
AR	0.46 (0.27)	0.24 (0.14)	0.50 (0.53)	0.46 (0.33)
SDCM	0.15 (0.11)	0.17 (0.17)	0.14 (0.10)	0.11 (0.05)
PDCM	0.08 (0.03)	0.08 (0.03)	0.08 (0.03)	0.08 (0.03)
LCB	0.32 (0.18)	0.48 (0.34)	0.62 (0.50)	0.42 (0.29)
HCB	0.23 (0.20)	0.48 (0.56)	0.86 (0.93)	0.45 (0.36)

Table 2.3: Average b_{bp} for within the productive zone for each biome

Average b_{bp} concentrations (m^{-1}) within the productive zone for each biome identified in SVD analysis. All values are calculated by interpolating profiles to 1-meter intervals and normalizing to depth of euphotic zone. Values are reported as the median (IQR). Note that in AR time seires, Winter, Spring, Summer and Fall correspond to Northeast Monsoon, Summer Intermonsoon, Southwest Monsoon, and Fall Intermonsoon, respectively.

	Depth integrated b_{bp} x 10^3			
	Winter	**Spring**	**Summer**	**Fall**
EQ	0.65 (0.14)	0.68 (0.13)	0.68 (0.15)	0.68 (0.16)
AR	0.94 (0.37)	0.78 (0.19)	1.14 (0.63)	0.94 (0.34)
SDCM	0.61 (0.38)	0.75 (0.45)	0.64 (0.31)	0.52 (0.18)
PDCM	0.45 (0.09)	0.48 (0.09)	0.50 (0.08)	0.47 (0.08)
LCB	0.85 (0.49)	1.40 (0.78)	1.57 (0.90)	1.06 (0.58)
HCB	0.87 (0.59)	1.38 (1.33)	1.86 (1.25)	1.10 (0.62)

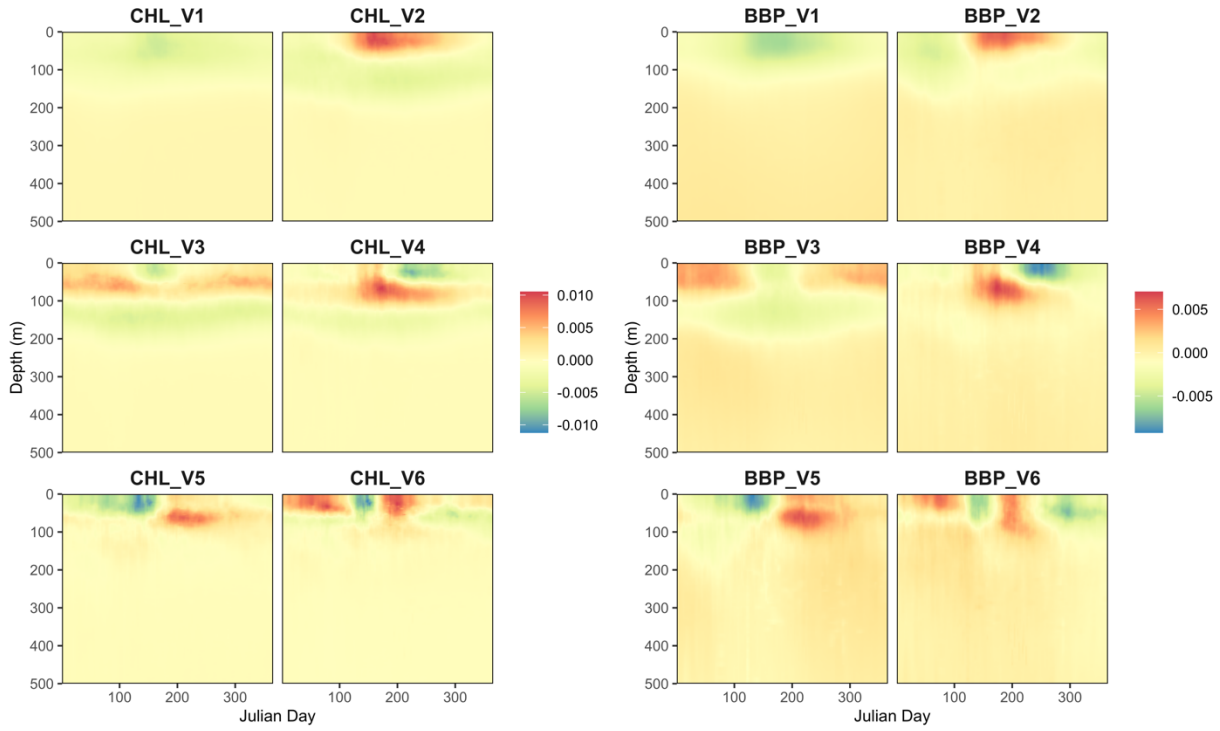

Figure 2.1: First six singular vectors obtained via SVD analysis

First six singular vectors obtained for rows of SVD matrix corresponding to measurements of [Chla] (CHL_V1 – CHL_V6) and bbp (BBP_V1 – BBP_V6). Although SVD was performed for measurements from 0 to 750 m, only depths up to 500 m are shown here. Units are arbitrary in either plot. Products of singular vectors and SVD coefficients yield scaled [Chla] and bbp values in original SVD matrix.

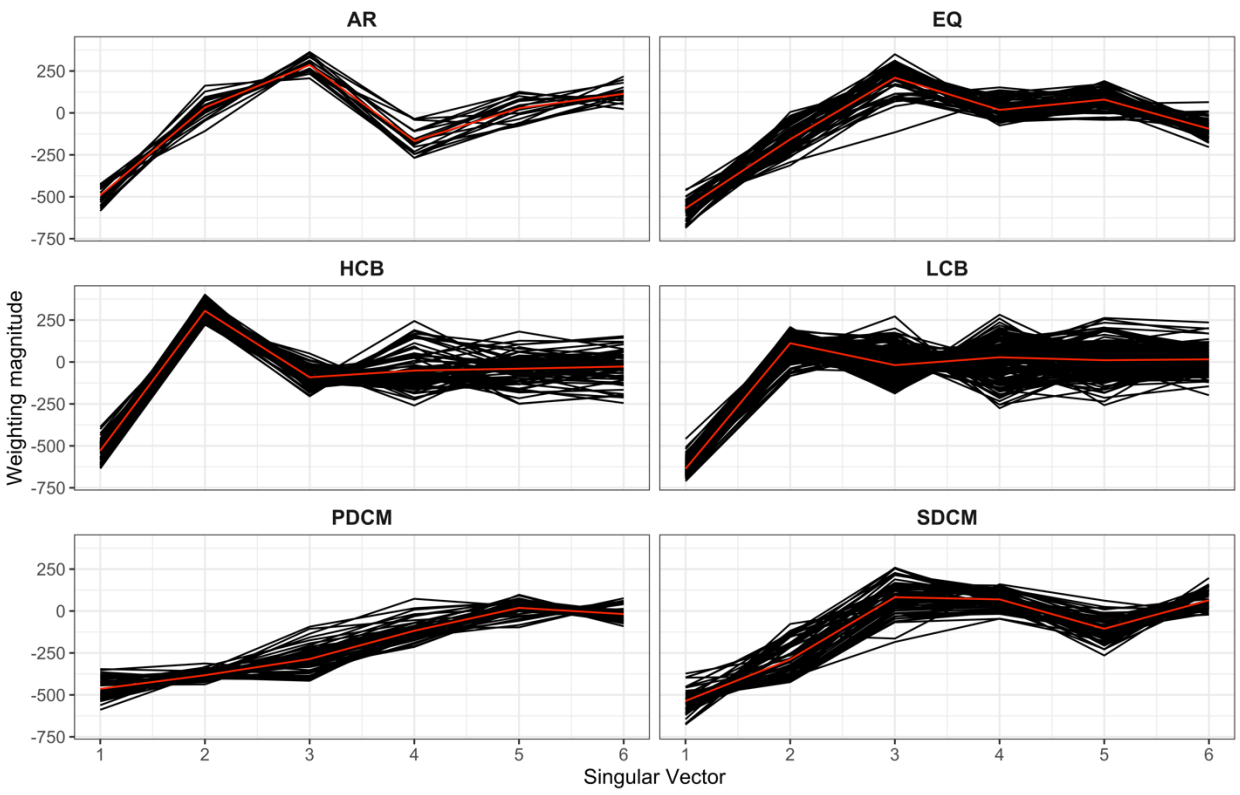

Figure 2.2: Coefficients for first six singular vectors obtained via SVD analysis

Weightings corresponding to the first six singular vectors obtained via singular value decomposition applied to the global BGC-Argo data set of [Chl*a*] and b_{bp} profiles, grouped by biomes identified via cluster analysis. Red line corresponds to the mean weighting magnitude for each vector

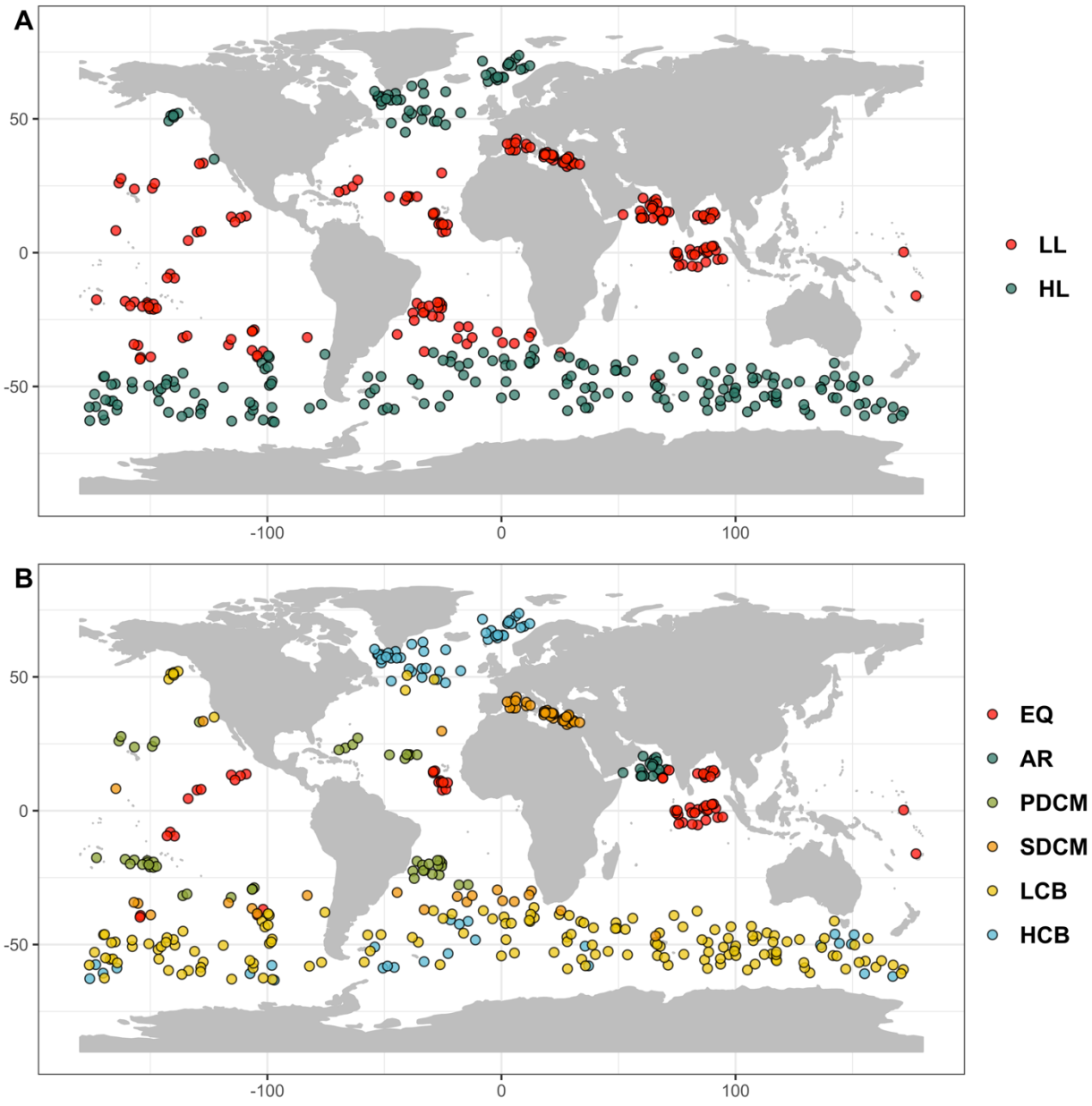

Figure 2.3: Geographic distribution of time series with biome classifications

Geographic distributions of **A)** High-latitude (HL) and low-latitude (LL) time series based on initial cluster analysis, after manual reassignment of Arabian Sea time series; **B)** Equatorial (EQ), Arabian Sea (AR), seasonal DCM (SDCM), permanent DCM (PDCM), low-chlorophyll bloom (LCB), and high-chlorophyll bloom (HCB) time series based on second round of cluster

analysis, as described in text. Coordinates correspond to the median latitude and longitude for each time series.

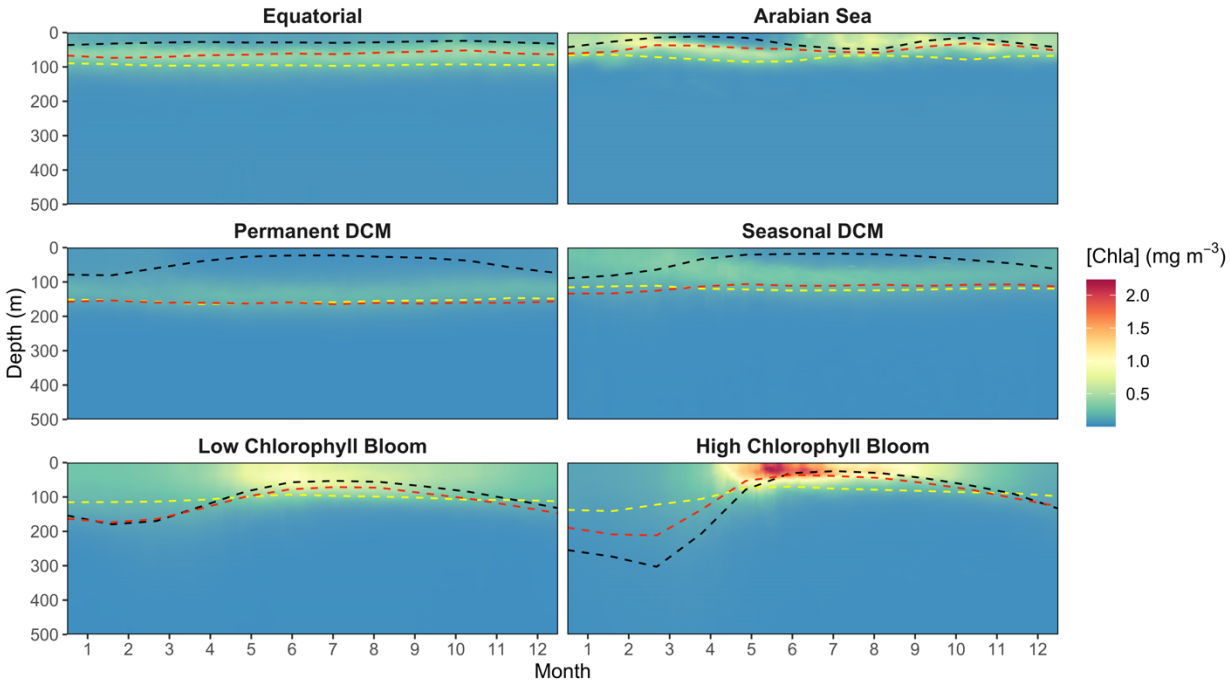

Figure 2.4: Average [Chl*a*] time series for each biome identified in cluster analysis

Mean [Chl*a*] (mg m^{-3}) for time series in the Equatorial (EQ), Arabian Sea (AR), Seasonal DCM (SDCM), Permanent DCM (PDCM), Low Chlorophyll Bloom (LCB), and High Chlorophyll Bloom (HCB) biomes. Black dashed line corresponds to mean mixed layer depth for each biome. Red dashed line corresponds to median nitracline depth for each biome. Yellow line corresponds to the median euphotic zone depth (defined as the 0.1% PAR isolume) for each biome.

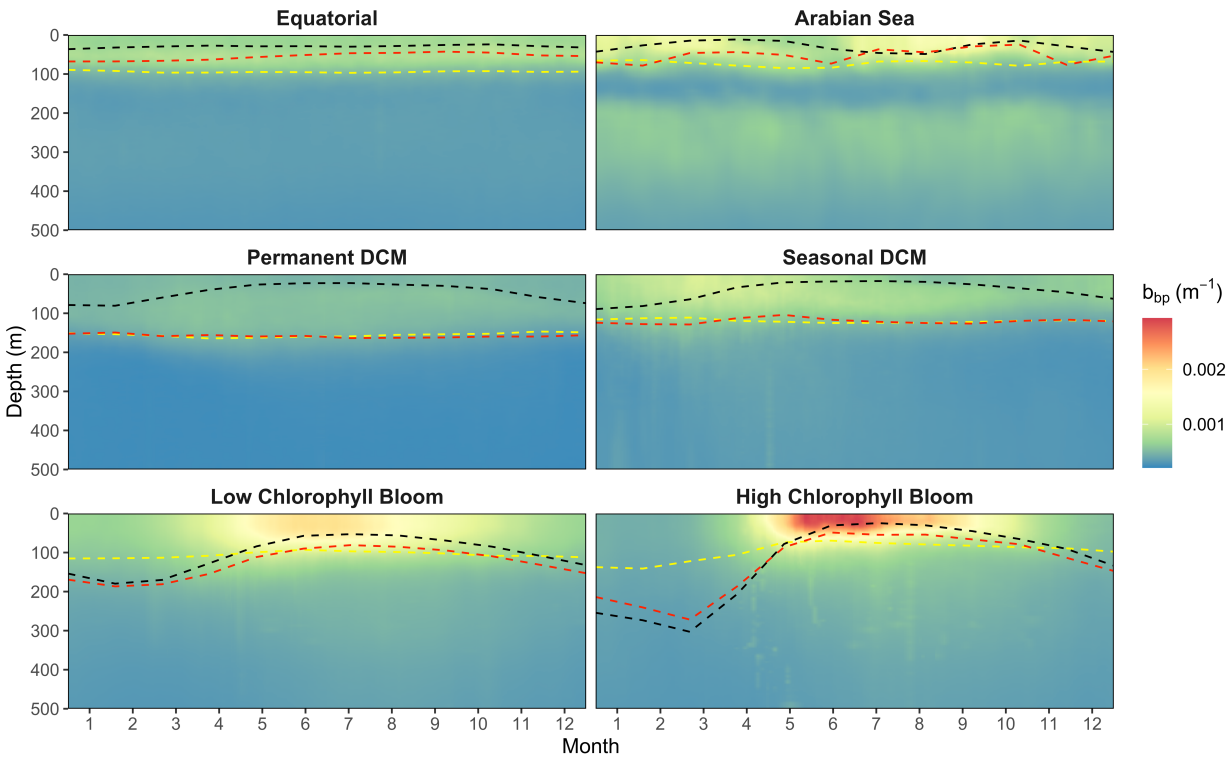

Figure 2.5: Average b_{bp} time series for each biome identified in cluster analysis

Mean b_{bp} (m^{-1}) for time series in the Equatorial (EQ), Arabian Sea (AR), Seasonal DCM (SDCM), Permanent DCM (PDCM), Low Chlorophyll Bloom (LCB), and High Chlorophyll Bloom (HCB) biome. Black dashed line corresponds to mean mixed layer depth for each biome. Red dashed line corresponds to median nitracline depth for each biome. Yellow line corresponds to the median euphotic zone depth (defined as the 0.1% PAR isolume) for each biome.

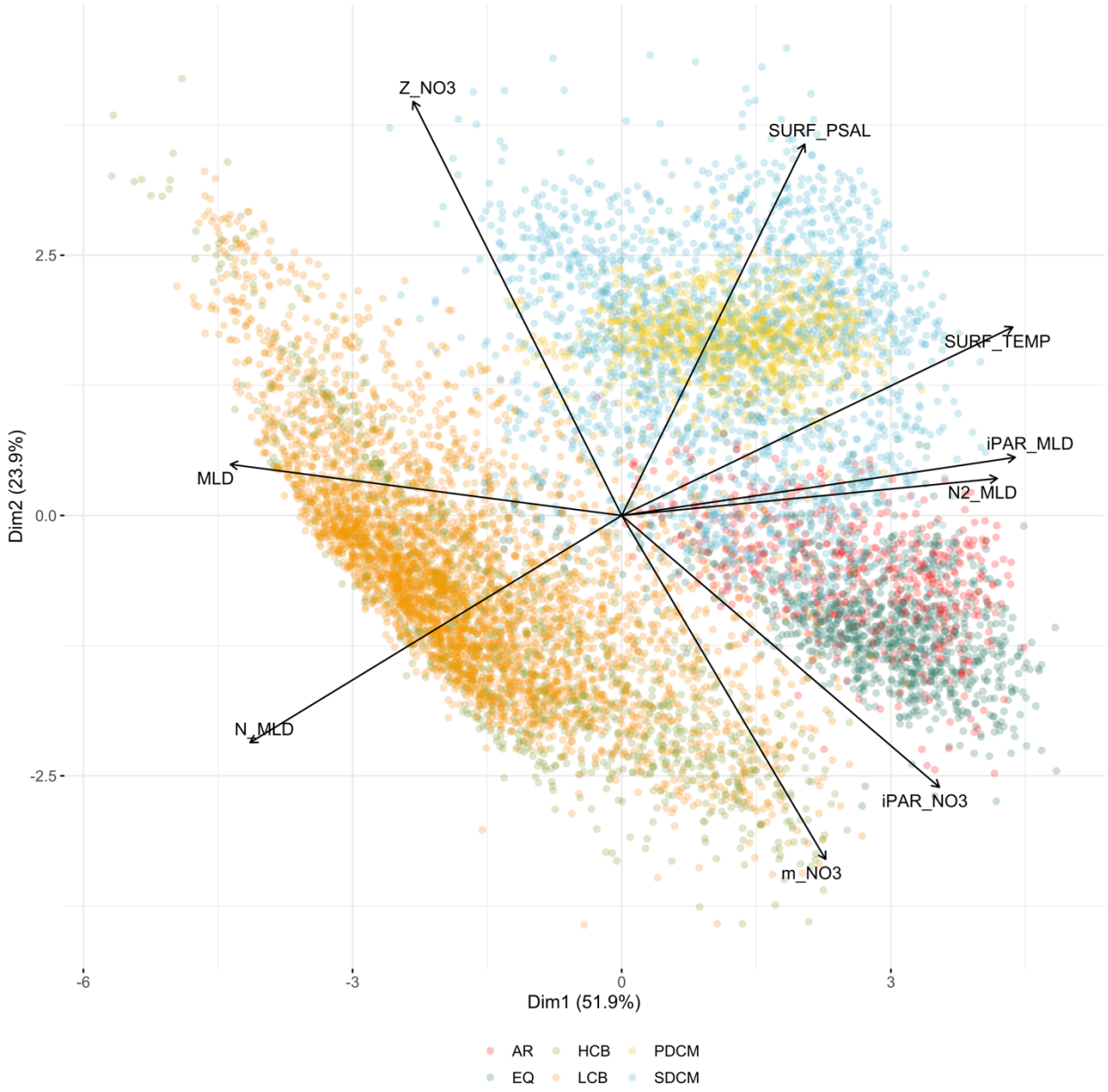

Figure 2.6: Results of principal component analysis of environmental parameters

Principal component analysis results for environmental parameters corresponding to individual profiles in Equatorial (EQ), Arabian Sea (AR), Seasonal DCM (SDCM), Permanent DCM (PDCM), Low Chlorophyll Bloom (LCB), and High Chlorophyll Bloom (HCB) biomes. Length of vectors for individual parameters corresponds to correlation between PCA loadings on each

component. Angle between vectors corresponds to correlation between PCA loadings for individual parameters. Parameters are the same as those in Table 2.1, using an underscore to indicate subscript text.

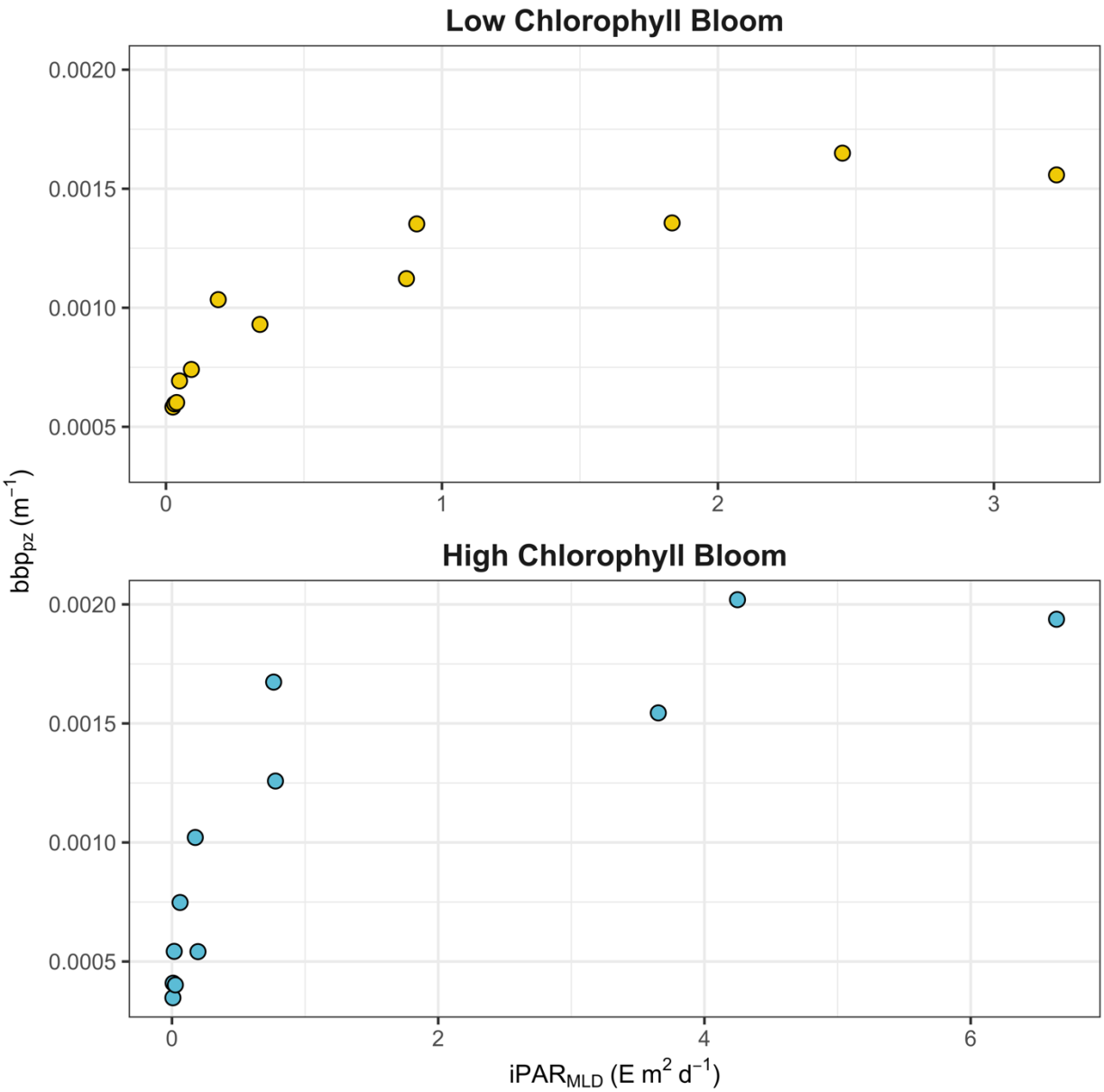

Figure 2.7: Relationships between nitracline depth and DCM parameters

Relationships between log-transformed values for $iPAR_{MLD}$ and bbp_{pz} for Low Chlorophyll Bloom (LCB), and High Chlorophyll Bloom (HCB) biomes. Individual points correspond to the average monthly value for each parameter. Note differences in x-axis scale for each plot.

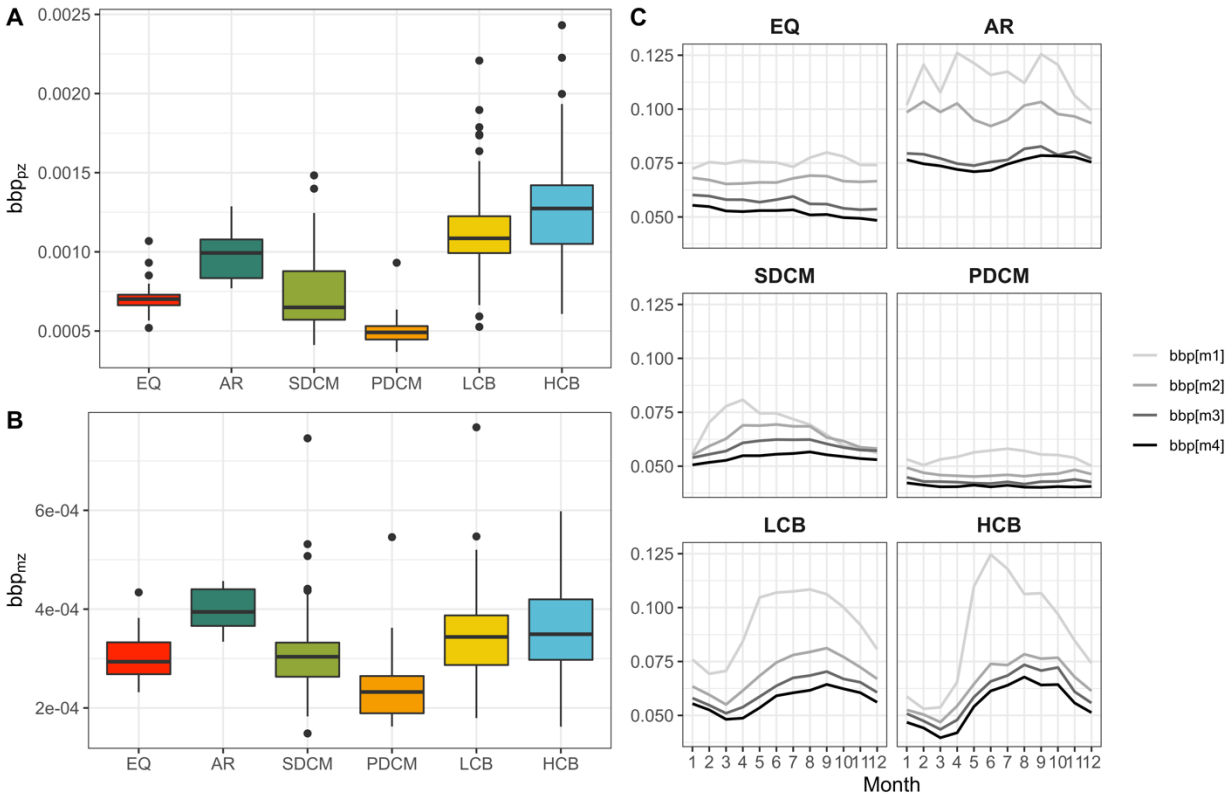

Figure 2.8: Median b_{bp} and trends in mesopelagic b_{bp} distributions for each biome

A) Distribution of bbp_{pz} values for Equatorial (EQ), Arabian Sea (AR), Seasonal DCM (SDCM), Permanent DCM (PDCM), Low Chlorophyll Bloom (LCB), and High Chlorophyll Bloom (HCB) biomes. **B)** Distribution of bbp_{mz} values for the same biomes. **C)** Average b_{bp} values within each of four equally-sized layers of the mesopelagic zone: bbp_{m1} – bbp_{m4}.

Chapter 3: A method for measuring per-cell chlorophyll *a* in natural picophytoplankton populations using fluorescence-activated cell sorting

3.1. Abstract

Marine phytoplankton play a central role in global biogeochemical cycling, carbon export, and marine ecosystem health. While chlorophyll *a* (Chl*a*) is widely used as a proxy for phytoplankton biomass, identifying the proportion of Chl*a* attributable to different phytoplankton groups remains a major challenge in oceanography, especially for the picophytoplankton groups that often represent the majority of phytoplankton biomass in the open ocean. This paper describes a method for measuring picophytoplankton per-cell Chl*a* in field samples using fluorescence-activated cell sorting followed by solvent based Chl*a* extraction and fluorescence quantification. Measurements of per-cell Chl*a* using this method are precise to within 2.8, 2.7, and 4.2 % for *Prochlorococcus*, *Synechococcus*, and picophytoeukaryotes (PPE), respectively. Applying this method to surface samples from the Gulf of Mexico, we determined per-cell Chl*a* to be 0.24 ± 0.07, 0.6 ± 0.33 and 26.36 ± 20.9 fg Chl*a* $cell^{-1}$ for *Prochlorococcus*, *Synechococcus*, and PPE, respectively. We demonstrate that this approach can be used to obtain estimates of group specific Chl*a* for *Prochlorococcus*, *Synechococcus*, and PPE, the latter two of which cannot be captured by existing methods. We also demonstrate that measurements of per-cell Chl*a* made using this method are sufficiently precise to capture relationships in field samples between per-cell Chl*a* and cytometer red fluorescence, providing a bridge from biomass estimates from cell counts and bulk measurements of total Chl*a*.

Chapter in preparation for publication; co-authors Ajit Subramaniam, Andrew Juhl, Joseph Montoya, and Solange Duhamel.

3.2. Introduction

Methods for evaluating marine phytoplankton community structure are essential for understanding biogeochemical cycling, the impacts of human activity on marine ecosystem composition and productivity, and marine food webs. Because measuring phytoplankton community structure over large spatial scales is labor intensive, numerous methods have been developed to infer phytoplankton community composition using cell size and/or phytoplankton accessory pigments. These approaches are generally based on size fractionated filtration or identifying diagnostic pigments expected to correlate with fractions of total chlorophyll *a* (Chl*a*) attributable to individual phytoplankton groups, as first proposed by Claustre (1994) and developed further by Vidussi et al. (2001). In the Vidussi et al. model (2001), the concentrations of seven pigments were identified as being diagnostic of Chl*a* fractions corresponding to three phytoplankton size classes: picophytoplankton, nanophytoplankton, and microphytoplankton, defined in the model as cells with a diameter < 2 μm, 2 – 20 μm, and > 20 μm, respectively. Refinements to this approach have used linear regression to relate diagnostic pigments to total Chl*a* concentrations (Uitz et al. 2006), introduced additional diagnostic pigments to differentiate *Prochlorococcus* from other pico-sized groups (Hirata et al. 2008, 2011), and have established criteria allowing for individual pigments to be diagnostic of multiple groups depending on total Chl*a* concentration (Brewin et al. 2010).

While estimates from diagnostic pigment-based methods converge with those derived from size fractionation in larger phytoplankton size classes (Brewin et al. 2014), the accuracy of pigment-based predictions for the three major picophytoplankton groups (*Prochlorococcus*, *Synechococcus*, and picophytoeukaryotes, hereafter PPE) is limited by the potential overlap of individual diagnostic pigments across multiple size classes (Uitz et al. 2009), and by the absence of a routinely measured pigment unique to *Synechococcus*. Depending on the diagnostic pigment

model used, this could lead to underestimates of the contributions to total Chl*a* by PPE in addition to ambiguity in the relative abundances of *Prochlorococcus* and *Synechococcus* in cases where both are present. Picophytoplankton generally dominate primary production in central gyres and in stratified surface waters. Because primary production estimates for *Prochlorococcus*, *Synechococcus*, PPE and larger phytoplankton vary by several orders of magnitude (Jardillier et al. 2010; Rii et al. 2016a; b; Duhamel et al. 2019), accurately resolving community structure within picophytoplankton is key to understanding marine primary production from local to global scales.

Because contributions of *Prochlorococcus*, *Synechococcus* and PPE populations to total Chl*a* cannot be disambiguated using current diagnostic pigment based or size fractionation based methods, we introduce a novel method for directly estimating picophytoplankton per-cell Chl*a* content (fg Chl*a* $cell^{-1}$) using fluorescence-activated cell sorting (FACS) in concert with solvent-based Chl*a* quantification. When combined with picophytoplankton cell counts and measurements of total Chl*a*, this method allows the estimation of picophytoplankton group specific Chl*a* concentration (µg Chl*a* ml^{-1}) and group specific Chl*a* fractions (% Chl*a*), thus bridging the two commonly used methods to describe phytoplankton biomass without the use of diagnostic pigments or other proxies. To demonstrate that this method is sufficiently sensitive to support these applications, we validated the approach using cultured phytoplankton and applied the approach to field samples from the Gulf of Mexico. Analyses of these field samples show that the method can also be used to investigate group specific relationships between per-cell Chl*a* and cytometer red fluorescence, as well as environmental variables, such as light.

Direct quantification of the picophytoplankton contribution to Chl*a* in field samples using this method could be used in the future to validate and improve quantification of phytoplankton

community structure based on diagnostic pigment methods. In addition, the method could be used, in combination with measurements of per-cell primary production (Jardillier et al. 2010; Rii et al. 2016a; b; Duhamel et al. 2019) or per-cell carbon content (Graff et al. 2012), to derive group specific photosynthetic parameters, or group specific chlorophyll:carbon ratios.

3.3. Materials and Procedures

3.3.1. Field sampling

Field samples were collected on research cruise EN642 in the Gulf of Mexico during July 2019 (Figure 3.1). Surface samples were collected from depths < 5 meters at thirteen sites. At five of these sites, additional samples were taken at depths corresponding to the halocline boundary and the deep chlorophyll maximum (DCM). All samples were collected using a Niskin bottle rosette and processed following the protocol described below. At all sampling sites, depth profiles of salinity, temperature and pressure were measured with a conductivity-temperature-depth instrument (CTD, Seabird 911 Plus) integrated within the bottle rosette. In situ irradiance for each sampling depth was measured using a PAR sensor on the rosette, and the percent of surface PAR (hereafter % I_0) was calculated by comparison to a shipboard PAR sensor.

For each sampling site/depth, separate subsamples were prepared for measurement of per-cell Chl*a*, HPLC pigments, and picophytoplankton cell abundance. To prepare per-cell Chl*a* samples, a peristaltic pump was used to concentrate cells from duplicate 1-liter samples onto 47-mm, 0.2-µm pore size polycarbonate membrane filters (Sigma Millipore catalog #GTTP04700). Each filter was folded and then placed in a cryovial containing 4 ml of seawater from the corresponding sampling depth. Samples were vortexed for 1 – 2 seconds to release cells from the filter, flash frozen in liquid nitrogen without the addition of biological fixatives and stored at -80° C. Water for HPLC analysis was collected from the same sampling depth as per-cell Chl*a* samples and filtered onto precombusted Whatman GF/F filters. The filters were then frozen in

liquid nitrogen until analysis at the NASA facility at Goddard Space Flight Center, following the method described in Van Heukelem & Thomas (2001). All filtrations were performed under subdued lighting conditions to prevent pigment degradation. For cell enumeration, 2-mL subsamples were fixed (1 % w/v electron microscopy grade paraformaldehyde, Electron Microscopy Sciences catalog RT-15710) for 10 – 15 minutes in the dark at room temperature. Samples were then flash-frozen in liquid nitrogen and stored at -80º C. All samples were processed within 8 months of collection. Only samples collected from depths corresponding to > 0.1 % I_0 and with salinity > 25.9 ppt were included in analyses.

3.3.2. Analytical processes for measurement of per-cell Chla

For all samples, a BD influx flow cytometer (BD Life Sciences, USA) was used to isolate cells from individual picophytoplankton populations by FACS. Cells from each population were collected onto separate 13-mm, 0.2-µm pore polycarbonate filters (Millipore Sigma catalog #GTTP01300) connected to a low-vacuum pump as described in Berthelot et al. (2019). A 70-µm nozzle was used, with sheath fluid pressure of 30 PSI (207 kPa) and sample fluid pressure of 31 PSI (214 kPa). At the start of each day of sample processing, the drop delay was calibrated using Accudrop Beads (BD Life Sciences, USA) and sorting efficiency was verified manually by sorting a specified number of 1-µm yellow–green microspheres (Fluoresbrite, Polysciences catalog #17154-10) onto a glass slide and counting the beads under an epifluorescence microscope. The breakoff position was monitored through the day and adjusted if necessary. The instrument was set at high sorting purity (1.0 drop mode). To maintain high sorting purity and recovery, samples were diluted in artificial seawater filtered through a 0.2-µm pore size filter as required to achieve an event rate of < 5,000 per second. Higher event rates resulted in

coincidence and swarm (e.g., coincidence leading to permanent scatter and fluorescence signals), as previously reported using the Influx (Kormelink et al. 2016).

To ensure consistent instrument alignment, the position of the forward scatter detector, 488nm laser, and 595nm laser were calibrated such that reference beads were aligned to a predetermined position on a calibration plot comparing forward scatter and red fluorescence after adjusting the instrument to clearly resolve *Prochlorococcus* populations in surface samples. Samples were additionally amended with fluorescent reference beads (1-µm yellow-green microspheres), to monitor alignment and focus of the instrument.

To gate populations of interest, cell doublets were first discarded using a pulse width versus forward scatter plot. That is, assuming that pulse width should generally covary with forward scatter for single particles (doublets have similar forward scatter as single cells, but increased pulse width), total events were gated so that only events with similar pulse width:forward scatter ratios were included in analyses. *Prochlorococcus*, *Synechococcus* and PPE were distinguished based on forward scatter (FSC), orange fluorescence (580/30 nm), and red fluorescence (>650 nm) as described in Bock et al. (2018). Briefly, particles with low forward scatter, low red fluorescence, and low orange fluorescence were identified as *Prochlorococcus* cells. Particles with intermediate forward scatter, intermediate red fluorescence, and high orange fluorescence were identified as *Synechococcus* cells. Particles with high forward scatter, high red fluorescence, and low orange fluorescence were identified as PPE cells. The BD Influx cytometer is equipped with a special forward scatter detector that increases the sensitivity of forward scatter photomultiplier tubes to small particles (e.g., up to 200 nm in diameter), making forward scatter a better discriminant for small cells like *Prochlorococcus* than side

scatter on this instrument. However, side scatter might be a more appropriate proxy for cell size on other instruments.

For each field sample, cells from each phytoplankton group were sorted to achieve a range of cell concentrations across three filters. Tests using culture samples determined > 3.0 x 10^5 cells per filter were sufficient to exceed fluorometer quantitation limits (determined as described below) for cyanobacteria, and that > 1.5 x 10^4 cells were sufficient to exceed fluorometer quantitation limits for PPE. Accordingly, cyanobacterial filters were prepared, at minimum, with 3 x 10^5, 5 x10^5, and 7 x 10^5 *Prochlorococcus* or *Synechococcus* cells per filter. PPE filters were prepared, at minimum, with 1.5 x 10^4, 2 x 10^4, and 2.5 x 10^4 cells per filter. Immediately after sorting and filtration, each filter was placed in a 2-ml screw-cap o-ring microcentrifuge tube (Fisher Scientific catalog #21-403-200) and stored at -80° C for at least one hour prior to Chl*a* extractions.

Chl*a* extractions were performed using 100% methanol following established methods (Strickland and Parsons 1972; Arar and Collins 1997). 100% methanol was chosen as an extraction solvent based on its reported superior extraction efficiency compared to other solvents (Holm-hansen and Riemann 1978; Wright et al. 1997). Fluorescence of Chl*a* in extracts was quantified using a Turner Trilogy fluorometer (Turner Designs, San Jose, CA, USA). To maximize the final Chl*a* concentration, the extraction volume was only 1.5 ml of 100% methanol. This volume was sufficient to both completely immerse the 13-mm diameter filters and to cover the measurement window of the Turner Trilogy fluorometer when using 12 mm (OD) borosilicate glass tubes (Fisher Scientific catalog # 14-961-26). Following the methanol addition, sample tubes were capped and vortexed in the dark for 15 seconds each, then extracted for 20 to 24 hours in the dark at -4° C. After the extraction period, samples were brought to room

temperature in the dark and vortexed for 15 seconds each to ensure homogenization. Finally, disposable transfer pipettes were used to transfer extracts to glass tubes for fluorescence quantification.

3.3.3. Determination of fluorometer detection and quantitation limits

To evaluate fluorometer detection and quantitation limits, a commercially available Chl*a* calibration standard (Sigma Aldrich catalog #479-61-8) was diluted in triplicate to concentrations ranging from 1 x 10^{-5} to 2 x 10^{-2} µg Chl*a* ml^{-1}. Chl*a* fluorescence of diluted samples was measured using a narrow bandpass filter set (excitation: 436/10 nm, emission: 685/10 nm). The narrow bandpass filter set lowers fluorescence signals produced by phaeopigments, thereby eliminating the need for acidifying samples during measurements (Welschmeyer 1994; Arar and Collins 1997). Although the use of a narrow bandpass filter set precluded the measurement of chlorophyll *b* or chlorophyll *c*, we found that the narrow bandpass filter yielded considerably lower variability in measurements of Chl*a* fluorescence compared to broadband filters (excitation ≤ 500 nm, emission ≥ 665 nm), especially at low Chl*a* concentrations (Appendix B.1; Figure B.1).

Fluorometer quantitation limits (x_L) were calculated as:

$$x_L = \bar{x}_{bi} + k s_{bi}$$

where $\bar{x}_{bi}$ is the mean of triplicate raw instrument values corresponding to a methanol blank, s_{bi} is the standard deviation of the blank measurements (IUPAC 1997). A coefficient of 10 was used as the coefficient *k*, a value commonly used in the determination of quantitation limits (MacDougall et al. 1980).

3.3.4. Comparison of cytometer red fluorescence to per-cell Chla in field samples

Cytometer red fluorescence is widely used as a proxy for intracellular Chl*a* and total phytoplankton Chl*a* for field samples (Veldhuis et al. 1997; Calvo-Díaz et al. 2008; Graff et al.

2016; Brewin et al. 2019). To assess whether the red fluorescence versus per-cell Chl*a* relationship was consistent across the three picophytoplankton groups studied in our samples, we compared per-cell Chl*a* and average group specific red fluorescence for picophytoplankton populations in cell abundance samples. To account for any adjustments to the sensitivity of red fluorescence detectors between samples, average group specific red fluorescence was normalized to the red fluorescence of fluorescent reference beads (1-µm yellow–green microspheres). Red fluorescence values were log-transformed for *Prochlorococcus* samples to linearize relationship with per-cell Chl*a*. Similar transformations did not improve regression fits for *Synechococcus* or PPE and so were not applied.

3.3.5. Calculation methods

Raw fluorescence output values (relative fluorescence units; hereafter RFU) measured on the Trilogy Fluorometer were converted to µg Chl*a* ml^{-1} using a calibration curve prepared with the same standard used for tests of instrument quantitation limits. To estimate per-cell Chl*a*, regressions were performed comparing the filter Chl*a* content with the corresponding number of cells sorted onto the filter (Figure 3.2). Average RFU values for triplicate measurements of methanol blanks were represented in regressions as filters containing zero cells. Regression slopes were interpreted as per-cell Chl*a* (fg Chl*a* $cell^{-1}$).

To evaluate sources of variability in estimates of per-cell Chl*a*, the ratio of sample standard error to sample means (hereafter relative error) was calculated for different steps of sample processing and multiplied by 100 (to express as a percentage). Three potential sources of variability in estimated per cell Chl*a* were evaluated: 1) To evaluate variability attributable to fluorometry measurements, relative error was calculated for triplicate measurements of calibration standards within the range of Chl*a* concentrations encountered when performing Chl*a* extractions, 2) To evaluate variability attributable to cell sorting and extraction protocols, relative

error of per-cell Chl*a* measurements was calculated as the ratio of regression slope standard error to regression slopes, and 3) To evaluate variability due to regressions, relative error for regression estimates of per-cell Chl*a* was compared relative to error for per-cell Chl*a* estimates obtained by normalizing filter Chl*a* content to the number of cells sorted per filter.

To estimate group specific Chl*a* concentration for *Prochlorococcus*, *Synechococcus*, and PPE, per-cell Chl*a* measurements for each sample were multiplied by the cell abundance of the corresponding phytoplankton group. Group specific Chl*a* fractions were calculated as the ratio between group specific Chl*a* and total Chl*a* (from the HPLC measurements) of each sample multiplied by 100 (to express as a percentage). For each field sample, total picophytoplankton Chl*a* was calculated as the sum of *Prochlorococcus*, *Synechococcus* and PPE Chl*a* fractions; prokaryote Chl*a* was calculated as the sum of measured *Prochlorococcus* and *Synechococcus* Chl*a* fractions; and nanophytoplankton + microphytoplankton Chl*a* was calculated as the difference between total Chl*a* and picophytoplankton Chl*a*. To compare field measurements of group specific Chl*a* fractions to those estimated by diagnostic pigment analysis (DPA) methods, diagnostic pigments quantified for each sample using HPLC were used to estimate group specific Chl*a* fractions following the approaches described in Vidussi et al. (2001), Uitz et al. (2006) and Hirata et al. (2011).

3.3.6. Statistical analysis

All data analysis was performed using R Studio (R Core Team 2016). All averaged values are reported ± the standard error of the mean. Analyses were performed on samples from individual depth bins (surface, halocline, or DCM), and on pooled data combining samples from all depth bins. Bivariate comparisons between estimated and measured group specific Chl*a* and HPLC pigment concentrations were performed using Pearson's correlation. ANOVA was used to identify significant differences in average parameter values between time points in storage

duration tests, and between measured and estimated group specific Chl*a* fractions. Tukey's honest significance difference post-hoc test was used to compare pairs of group means when ANOVA indicated significant between-group differences.

Linear regressions were used to evaluate relationships between per-cell Chl*a* and cytometer red fluorescence, and to evaluate the change in per-cell Chl*a* with changes in % I_0. All regression analyses were performed using original least squares. Statistical tests of regression fits are based on F-tests evaluating the ratio of model residuals to differences between regression estimates and mean values of predictor variables. A significance threshold of 0.05 was used for all statistical tests.

3.4. Results

3.4.1. Measurement of per-cell Chla in field samples

Samples were collected from stratified waters across a salinity gradient ranging from 25.9 to 36.2 ppt, and with surface temperatures ranging from 23.8 to 29.6° C. Total Chl*a* concentrations across all sampling depths ranged from 0.12 to 3.22 mg m^{-3}. Regressions used to calculate per-cell Chl*a* were statistically significant for all phytoplankton groups across all sampling sites ($R^2 > 0.97$, $p < 0.05$ for all comparisons). Regression fits were not affected by the duration of cell sorts. Mean estimates of per cell Chl*a* for *Prochlorococcus*, *Synechococcus*, and PPE for different depth bins are shown in Table 3.1. Across all field samples, the relative error of per-cell Chl*a* measurements was 2.87 %, 2.76 %, and 4.19 % for *Prochlorococcus*, *Synechococcus*, and PPE, respectively (Table B.1). *Prochlorococcus* and *Synechococcus* per-cell Chl*a* increased 3-fold between the surface and the deep chlorophyll maximum (DCM), while PPE per-cell Chl*a* increased 2-fold between the surface and the DCM (Table 3.1). Differences in per-cell Chl*a* between depth bins were statistically significant for *Prochlorococcus* (ANOVA, $F_{2,15} = 13.9$; $p = 0.004$) and *Synechococcus* (ANOVA, $F_{2,21} = 7.77$; $p = 1 \times 10^{-4}$), but not for PPE (ANOVA, $F_{2,18}$

= 3.34; p = 0.09). Regressions between per-cell Chl*a* and cytometer red fluorescence were statistically significant for all phytoplankton groups (Figure 3.3; ANOVA, $p < 0.05$ for all comparisons).

Mean estimates of group specific Chl*a* fractions are shown in Table 3.2. When averaging across all sampling sites, there were no significant differences between measured total picophytoplankton Chl*a* fractions and those estimated based on concentrations of HPLC pigments (Table 3.2; ANOVA; $p > 0.05$ for all comparisons). However, no statistically significant relationships were observed in comparisons of estimated Chl*a* fractions when comparing individual samples where both DPA and FACS methods were applied, with the exception of *Prochlorococcus* Chl*a* fractions, which were significantly correlated with *Prochlorococcus* Chl*a* fractions estimated using the Hirata et al. (2011) model (Pearson's r = 0.94; $p = 1 \times 10^{-5}$). Group specific Chl*a* concentrations (mg Chl*a* m^{-3}) measured by FACS generally correlated with their corresponding diagnostic pigments (Table 3.3). There were no significant correlations between group specific Chl*a* fractions (% Chl*a*) and concentrations of HPLC pigments, with the exception of divinyl Chl*a* (DV-Chl*a*), which was significantly correlated with measured *Prochlorococcus* Chl*a* fractions (Pearson's r = 0.78; p = 0.004).

Regressions comparing per-cell Chl*a* to $\log_{10}$ transformed % I_0 were statistically significant for all phytoplankton groups (Table 3.4; Figure 3.4). Based on regression R^2 values, changes in % I_0 accounted for 62 %, 34 %, and 53 % of variability in per-cell Chl*a* for *Prochlorococcus*, *Synechococcus*, and PPE, respectively. Because % I_0 depths were $\log_{10}$ transformed, regression slopes can be interpreted as the change in per-cell Chl*a* for an order of magnitude change in % I_0. Therefore, regressions of per-cell Chl*a* to log-transformed % I_0

showed per-cell Chl*a* to increase by 0.26 fg, 0.40 fg and 18.65 fg for *Prochlorococcus*, *Synechococcus*, and PPE, respectively, for an order of magnitude decrease in % I_0.

3.5. Discussion

Our results demonstrate that FACS followed by Chl*a* extraction and quantification can be used to measure picophytoplankton per-cell Chl*a* in field samples, yielding results that are similar to previously reported values from culture studies (Table 3.5) and providing critical insight into *in situ* phytoplankton community structure that is not available using existing methods. In combination with independently measured picophytoplankton cell counts and total Chl*a* concentration, the FACS method uniquely provides estimates of Chl*a* fractions for *Synechococcus* and PPE, picophytoplankton groups not individually represented in DPA methods. At the same time, the FACS method provides estimates of *Prochlorococcus* Chl*a* that overlap those provided by DPA methods. Similarly, summed group specific Chl*a* fractions measured using the FACS method converge with estimates of total picophytoplankton Chl*a* fractions provided by DPA methods. Thus, this new method is compatible with earlier approaches to quantifying phytoplankton community structure where there is overlap in quantitation, while also providing unique insights.

3.5.1. Comparison of estimated and measured group specific Chla fractions

Convergence of measured and estimated *Prochlorococcus* Chl*a* are attributable to the strong correlations between measured *Prochlorococcus* Chl*a* and DV-Chl*a*, which is the only diagnostic pigment used for *Prochlorococcus* in the Hirata et al. (2011) approach. Weak relationships between estimated and measured group specific Chl*a* fractions for comparisons of individual sampling sites are likely primarily attributable to the fact that HPLC pigments in our data were much more strongly correlated with absolute group specific Chl*a* (mg Chl*a* m^{-3}) than with group specific Chl*a* fractions (% total Chl*a*), even after normalizing HPLC pigments to the

sum of the individual diagnostic pigments, as is typically performed in diagnostic pigment models. This being the case, measured group specific Chl*a* fractions would not be expected to covary with those estimated from HPLC pigments in our data. It also may be the case that the combination of pigments used in the different DPA methods are not applicable to the region sampled. For example, we found Chl*b*, 19-hex fucoxanthin, and 19-but fucoxanthin to be the best predictors of measured picophytoplankton Chl*a* in our data, as opposed to zeaxanthin and Chl*b*, as used in the Vidussi et al. (2001) and Uitz et al. (2006) approaches; or zeaxanthin alone, as used in the Hirata et al. (2011) model. These challenges in identifying region-specific diagnostic pigment and/or weightings underline the issues raised in the introduction and further demonstrates the usefulness of the FACS-based approach.

3.5.2. Relationships between per-cell Chla and cytometer fluorescence

The high R^2 values obtained for regressions between cytometer red fluorescence and measurements of per-cell Chl*a* support the use of cytometer red fluorescence as a proxy for per-cell Chl*a* and suggest that results obtained using this method could be used to determine instrument and/or region specific "calibration coefficients" relating per-cell Chl*a* to cytometer red fluorescence. That is, this method could be used to measure per-cell Chl*a* for a subset of field samples, and the relationship between cytometer red fluorescence then used to estimate per-cell Chl*a* for samples where cytometry cell counts are available alone. Our results also demonstrate the potential for differing relationships between cytometer red fluorescence and per-cell Chl*a* for individual phytoplankton groups (Figure 3.4). This underlines the importance for accounting for such group specific differences when using cytometer red fluorescence as a proxy for per-cell Chl*a*. Similar group specific differences in fluorescence properties have been reported previously in culture studies, and have been ascribed in part to differences in pigment composition among taxa (Sosik et al. 1989; Morel et al. 1993). Relationships between red fluorescence and per-cell

Chl*a* may also be affected by cell size, as reflected in equations derived from regressions comparing red fluorescence to per-cell Chl*a*. Assuming *Prochlorococcus* to have the smallest cell diameter of the groups studied and PPE to be the largest, and taking any transformations into account, regression coefficients reflect a decrease in the sensitivity of red fluorescence to changes in per-cell Chl*a* from smaller to larger groups. Similar non-linear relationships between fluorescence and per-cell Chl*a* with changes in cell size have been reported in culture studies (Álvarez et al. 2017), and presumably reflect reductions in absorption efficiency with increases in cell size and/or pigment density (Morel and Bricaud 1981).

3.5.3. Relationships between per-cell Chla and irradiance

The observed increases in per-cell Chl*a* in response to reductions in % I_0 (Table 3.4; Figure 3.4) are similar to those previously reported in culture experiments investigating change in per-cell Chl*a* with irradiance for both cyanobacterial groups (Moore et al. 1995) and small eukaryotes (Sosik et al. 1989). That relationships between *Prochlorococcus* per-cell Chl*a* and % I_0 were stronger than for other groups is also consistent with previous studies demonstrating *Prochlorococcus* to adjust cellular chlorophyll concentrations over a wide range of irradiances, especially relative to *Synechococcus* (Moore et al. 1995; Marie et al. 1999). That % I_0 accounts for at most 62 % total variance in per-cell Chl*a* (depending on group) indicates per-cell Chl*a* in picophytoplankton to be strongly influenced by other in situ variables, likely including growth rates (Blanchot et al. 1997; Neveux et al. 2003), cell size (Calvo-Díaz et al. 2008; Álvarez et al. 2017), and taxonomic composition (Veldhuis and Kraay 1993; Blanchot et al. 2001).

3.5.4. Comments and recommendations

Tested sources of variability in per-cell Chl*a* measurements were small relative to detection limits and variability in *in situ* per-cell Chl*a*. Most variability encountered in estimates of per-cell Chl*a* was attributable to fluorometry measurements or linear regressions (Table B.1).

Because the relative error of fluorometry measurements scales inversely with Chl*a* concentration (e.g., relative error was lower at higher Chl*a* concentrations), variability in estimates of per-cell Chl*a* for all groups may be lowered by sorting a larger number of cells per filter, if possible. The elevated relative error encountered for PPE is presumably due to the lower number of cells typically sorted for PPE filters compared to *Prochlorococcus* and *Synechococcus* filters. While sorting larger numbers of PPE cells is not likely practical in most cases, variability in estimates of per-cell Chl*a* for PPE may be lowered by preparing triplicate samples of cells sorted to provide an intermediate number of cells per filter and normalizing by cell abundance rather than by performing regressions.

We found that sample storage and extraction times did not significantly contribute to variability in measurements of per cell Chl*a* within the time frames tested (Appendix B.2 and B.3). However, the decrease in the number of intact picophytoplankton cells over 8 months storage (Figure B.2) demonstrates the importance of processing samples as soon after collection as possible for efficient sorting. While studies have shown that biological fixatives decrease cell loss during storage of picophytoplankton samples (Vaulot et al. 1989; Sato et al. 2006), we did not test the effect of fixation on per-cell Chl*a*. Given that strong relationships were identified between measured per-cell Chl*a*, HPLC pigment concentrations, and cytometry relative fluorescence across independent samples, despite processing per-cell Chl*a* measurements over a 2 – 8 month period following sample collection, the method is quite robust with regard to post-collection storage.

The method as described is optimized for sorting pico-sized cells. However, it could be adapted to measure per-cell Chl*a* for larger size classes as well, albeit with adjustments to sorting configurations. Tests of this method using laboratory cultures of the nano-sized chlorophyte

Pyramimonas parkeae indicated that as few as 200 cells per filter were sufficient to exceed fluorometer quantitation limits, suggesting that sorting sufficient numbers of these cells from natural samples may be realistic, even in areas with low total Chl*a* concentration. However, the sources of variability encountered in estimates of per-cell Chl*a* for PPE may be compounded in larger size classes, owing to the lower number of cells from these groups that can be sorted when processing samples.

Finally, in samples where the phytoplankton groups of interest are in low abundance (as was occasionally the case in this study for *Prochlorococcus* in coastal samples, or for *Synechococcus* in deep samples) the amount of time required to sort the minimum number of cells necessary to exceed fluorometer quantitation limits may become impractical. While it was generally feasible to completely process surface samples within 30 minutes each, this increased to as long as two hours for samples from the bottom of the euphotic zone. However, such increases in sort duration were not observed to affect regression fits in corresponding measurements of per-cell Chl*a*.

3.6. Conclusion

FACS combined with Chl*a* extraction can be used to quantify picophytoplankton per-cell Chl*a* with enough precision to support research applications in the field. Combined with picophytoplankton cell counts, this method provides a novel and practical approach for estimating fractions of total Chl*a* attributable to different picophytoplankton groups and for evaluating group specific variability in per-cell Chl*a* in response to environmental conditions. This new approach performs similarly to existing DPA methods for quantifying Chl*a* attributable to *Prochlorococcus* in field samples and is currently the only method that provides direct assessment of *Synechococcus* and PPE biomass in field samples. Our field results also validate

the use of cytometer red fluorescence as a proxy for per-cell Chl*a*. Given the widespread importance of picophytoplankton and the fundamental nature of per-cell Chl*a* to many oceanographic measurements, this method will be useful in a broad range of biological and biogeochemical studies.

Table 3.1: Summary of population characteristics

Values for per-cell Chl*a* (fg Chl*a* cell^{-1}), cell abundance (1 x 10^{10} cells m^{-3}), group specific Chl*a* (mg Chl*a* m^{-3}), and group contributions to total Chl*a* (%) for *Prochlorococcus* (PRO), *Synechococcus* (SYN), and picophytoeukaryotes (PPE) in field samples grouped by depth bin.

Group	Depth Bin	n	Per-cell Chl*a*	Cell abundance	Group specific Chl*a*	Contribution to total Chl*a*
PRO	Surface	5	0.24 ± 0.07	4.58 ± 4.22	0.01 ± 0.01	3.09 ± 1.69
	Halocline	7	0.55 ± 0.23	2.03 ± 3.99	0.04 ± 0.05	3.51 ± 7.06
	DCM	6	0.80 ± 0.21	4.96 ± 5.80	0.01 ± 0.01	9.88 ± 13.48
SYN	Surface	9	0.60 ± 0.33	32.09 ± 41.1	0.12 ± 0.11	15.83 ± 16.70
	Halocline	7	0.84 ± 0.33	15.3 ± 14.3	0.03 ± 0.03	17.97 ± 10.69
	DCM	6	1.51 ± 0.73	2.46 ± 2.64	0.15 ± 0.15	5.90 ± 3.69
PPE	Surface	7	26.36 ± 20.9	0.50 ± 0.64	0.07 ± 0.05	6.41 ± 2.44
	Halocline	7	39.82 ± 15.9	0.19 ± 0.12	0.10 ± 0.08	19.22 ± 9.94
	DCM	6	56.07 ± 26.9	0.19 ± 0.18	0.10 ± 0.11	20.22 ± 13.06

Table 3.2: Comparison of cytometric and HPLC estimates of Chl fractions

Group	This study	Vidussi et al. (2000)	Uitz et al. (2006)	Hirata et al. (2011)
PICO	28.2 ± 15.3	41.83 ± 19.0	37.9 ± 16.6	24.7 ± 17.6
Prok. (PRO + SYN)	15.6 ± 9.7	--	--	24.7 ± 17.6
PRO	7.3 ± 9.7	--	--	8.7 ± 9.9
SYN	9.4 ± 7.6	--	--	--
PPE	12.5 ± 8.9	--	--	--
NANO + MICRO	71.9 ± 19.0	50.8 ± 10.5	61.1 ± 15.8	73.7 ± 16.8

Table 3.3: Relationships between group Chl*a* and HPLC pigments

	Divinyl Chl*a*	Chl*b*	Fucoxanthin	Peridinin	19-hex-fucoxanthin	19-but-fucoxanthin	Alloxanthin	Zeaxanthin
PRO	**0.96****	**0.74****	--	--	0.70*	0.81**	--	--
SYN	--	--	--	0.88**	--	--	0.67*	**0.86****
PPE	--	**0.73****	--	--	**0.78****	**0.67***	0.58*	--
PICO	--	**0.83****	--	0.67*	**0.84****	**0.75****	--	--
NANO + MICRO	--	--	**0.92****	0.67*	--	--	**0.77****	--

Table 3.4: Relationships between per-cell Chl*a* and % I_0

Group	Intercept	Slope	R^2	p-value
PRO	0.73 ± 0.06	-0.26 ± 0.05	0.62	0.0003
SYN	1.26 ± 0.16	-0.40 ± 0.13	0.34	0.006
PPE	50.34 ± 5.14	-18.65 ± 4.27	0.53	0.0005

Table 3.5: Estimates of per-cell Chl*a* from this and previous studies

Source	Basin	Method	PRO	SYN	PPE
Estimates from field samples					
Blanchot et al. (1996)	SW Pacific	Size-fractionated samples normalized by cell abundance	0.2–6.4	NA	NA
Blanchot et al. (2001)	Eq. Pacific	Size-fractionated samples normalized by cell abundance	1.0–3.6	NA	15–187
Brunet et al. (2006)	Med.	Size-fractionated samples regressed by red fluorescence	NA	NA	17–168
Glover et al. (1988)	Sargasso Sea	Size-fractionated samples normalized by cell abundance	NA	0.50–7.50	NA
Murrell et al. (2004)	Gulf of Mexico	Size-fractionated samples regressed by cell abundance	NA	3.00–3.40	NA
This study	**Gulf of Mexico**	**FACS and direct quantification**	**0.21–1.43**	**0.31–3.41**	**7.59–139**
Estimates from culture samples					
Gustavs et al. (2009)	NA	Culture samples of *Synechocystis sp.* regressed by cell abundance	NA	NA	39 (22)
Kana & Gilbert (1987)	NA	Culture samples normalized by cell abundance	NA	1.0–5.0 (1330–30)	NA
Moore et al. (1995)	NA	Culture samples normalized by cell abundance	0.29–3.52 (110–3)	3.18–27.0 (462–5)	NA
Morel et al. (1993)	NA	Culture samples normalized by cell abundance	0.91–2.62 (100 – 25)	0.90–1.25 (100 – 25)	NA
Six et al. (2008)	NA	Culture samples of *Ostreococcus sp.* and *O. tauri* normalized by cell abundance	NA	NA	20–40 (800–20)
Stramski et al. (1995)	NA	Culture samples normalized by cell abundance	NA	4.00–6.90 (338–0)	NA
This study	**NA**	**FACS and direct quantification**	**0.85±0.03 (80)**	**1.48±0.03 (80)**	**10.44±0.39 (80)**

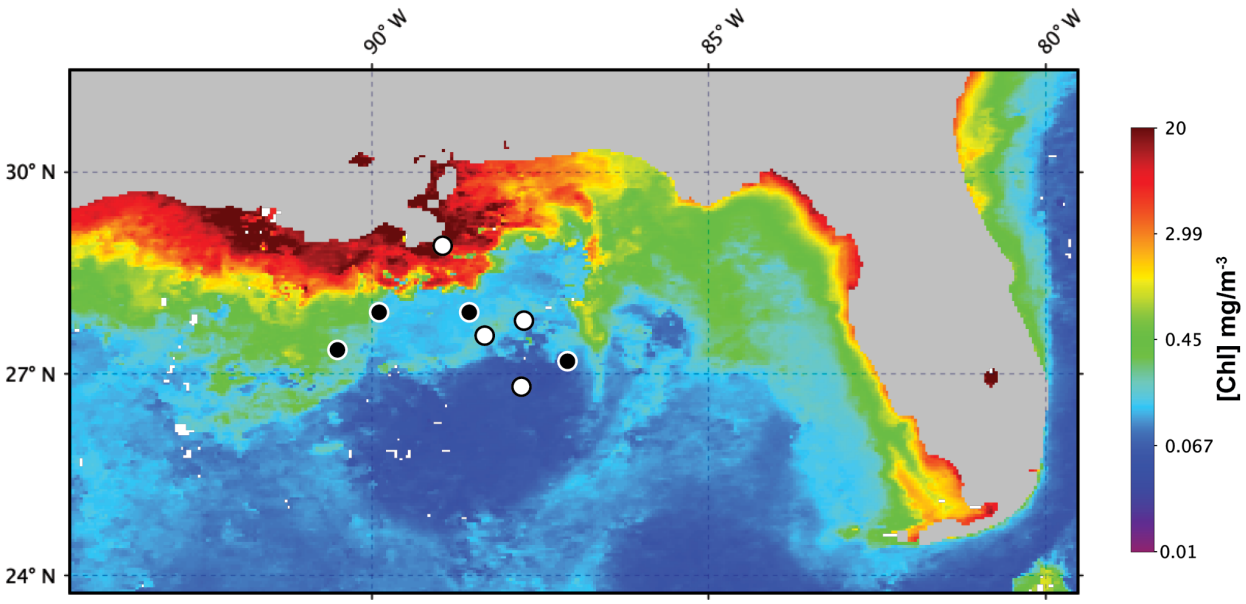

Figure 3.1: Sampling locations for EN642

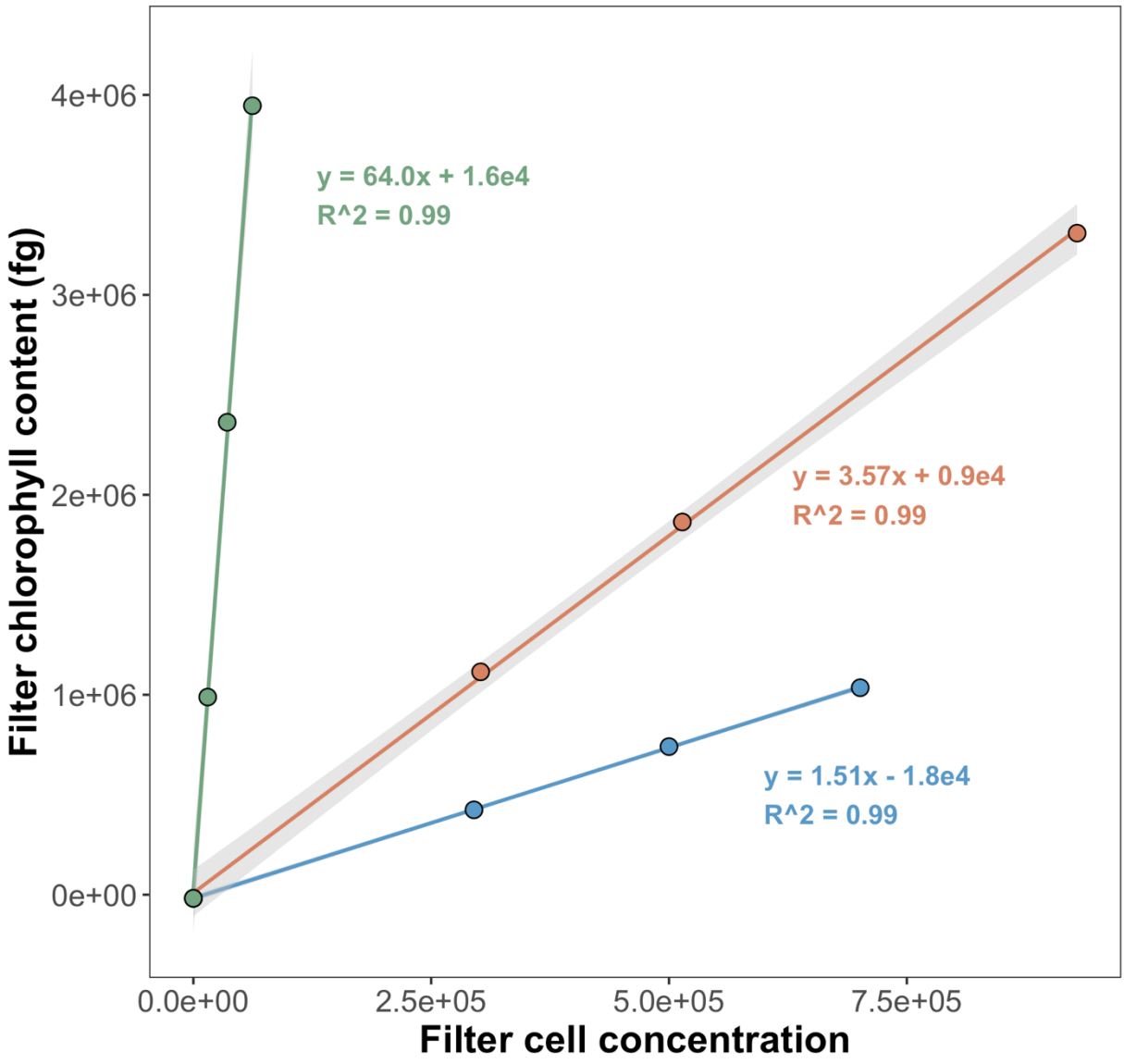

Figure 3.2: Regressions used for determination of per-cell Chl*a* based on cell sorts

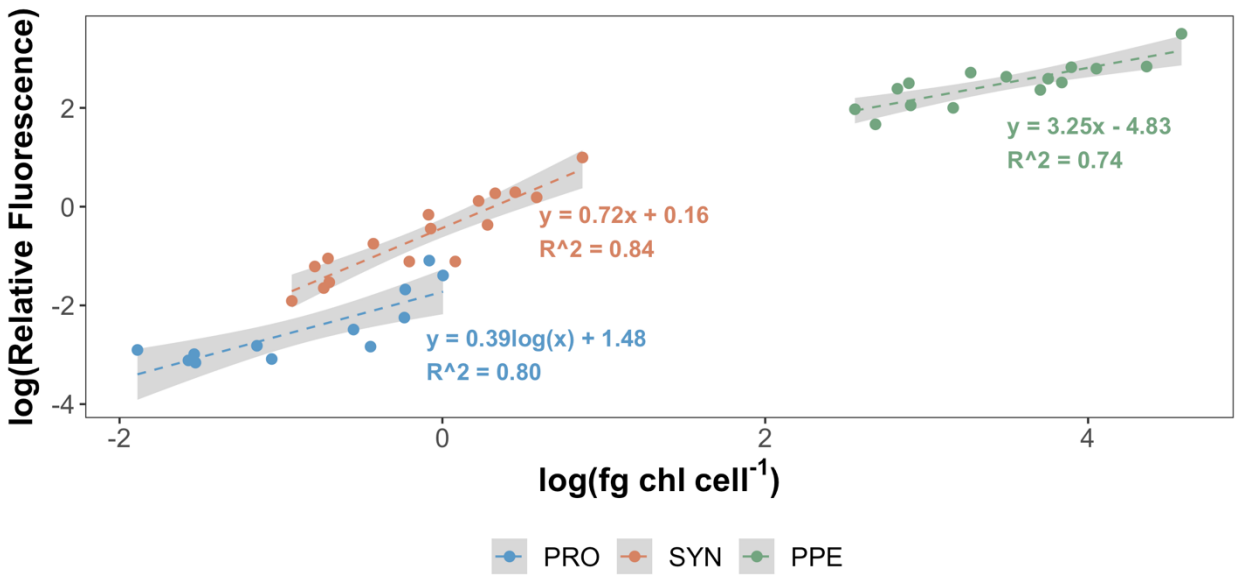

Figure 3.3: Comparison of relative fluorescence and cytometrically determined per-cell Chl*a*

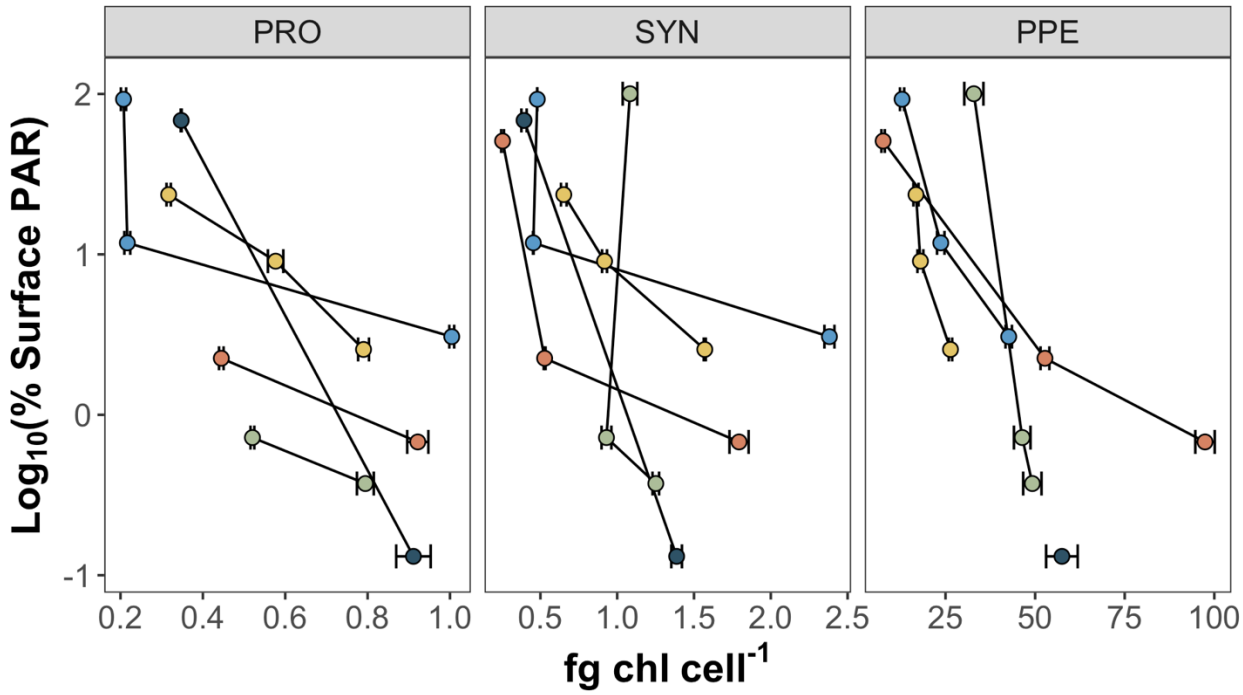

Figure 3.4: Relationship between $\%I_0$ and per-cell Chl*a*

Chapter 4: Microbial community structure in the Western Tropical South Pacific

4.1. Abstract:

Accounting for 40 percent of the earth's surface, oligotrophic regions play an important role in global biogeochemical cycles, with microbial communities in these areas representing an important term in global carbon budgets. While the general structure of microbial communities has been well documented in the global ocean, some remote regions such as the Western Tropical South Pacific (WTSP) remain fundamentally unexplored. Moreover, the biotic and abiotic factors constraining microbial abundances and distribution remain not-well resolved. In this study, we quantified the spatial (vertical and horizontal) distribution of major microbial plankton groups along a transect through the WTSP during the austral summer of 2015, capturing important autotrophic and heterotrophic assemblages including cytometrically determined abundances of non-pigmented protists (also called flagellates). Using environmental parameters (e.g., nutrients and light availability) as well as statistical analyses, we estimated the role of bottom-up and top-down controls in constraining the structure of the WTSP microbial communities in bio- geochemically distinct regions. At the most general level, we found a "typical tropical structure," characterized by a shallow mixed layer, a clear deep chlorophyll maximum at all sampling sites, and a deep nitracline. *Prochlorococcus* was especially abundant along the transect, accounting for 68 ± 10.6 % of depth-integrated phytoplankton biomass. Despite their relatively low abundances, picophytoeukaryotes (PPE) accounted for up to 26 ± 11.6 % of depth-integrated phytoplankton biomass, while *Synechococcus* accounted for only 6 ± 6.9 %. Our results show that the microbial community structure of the WTSP is typical of highly stratified regions and underline the significant contribution to total biomass by PPE populations. Strong relationships between N_2 fixation rates and plankton abundances demonstrate the central

Originally published as Bock, N., Van Wambecke, F., Dion, M., Duhamel, S. 2018. Microbial community structure in the Western Tropical South Pacific. *Biogeosciences*. DOI: 10.5194/bg-15-3909-2018

role of N_2 fixation in regulating ecosystem processes in the WTSP, while comparative analyses of abundance data suggest microbial community structure to be increasingly regulated by bottom-up processes under nutrient limitation, possibly in response to shifts in abundances of high nucleic acid bacteria (HNA).

4.2. Introduction

Subtropical oligotrophic oceans, covering approximately 40 percent of the earth's surface, represent one of the earth's largest biomes (Sarmiento et al. 2004). Despite reduced productivity compared to coastal regions, these nutrient-limited ecosystems are estimated to account for one-quarter of annual net marine primary production (Field et al. 1998). Therefore, it is of central importance to understand the factors shaping microbial communities in these regions, and to account for how these factors might vary seasonally and geographically between the ocean's major oligotrophic regions. There has been an enormous deal of progress toward this goal over the last three decades. The groups of phytoplankton that numerically dominate the open ocean—*Prochlorococcus*, *Synechococcus*, and picophytoeukaryotes (PPE)—are of small size, generally < 2 μm, with the most abundant group, *Prochlorococcus*, identified in the late 1980s (Chisholm et al. 1988). Since then, the widespread use of flow cytometry to characterize microbial communities has led to the publication of numerous studies documenting the distribution of these organisms (Veldhuis and Kraay 2000). Especially over the last ten years, molecular methodologies have allowed for the characterization of these groups at the taxon or species level, revealing enormous diversity across all trophic levels of marine microbial communities and identifying numerous ecotypes occupying distinct ecological niches (Venter et al. 2004; Carlson et al. 2007). More recently, environmental sequencing has revealed a surprising diversity of small sized PPE and new eukaryotic lineages continue to be discovered and characterized (Lepère et al. 2009; Kashtan et al. 2014; Kim et al. 2016; Rii et al. 2016a), while also revealing the importance of viruses in regulating phytoplankton communities (Brum et al. 2015; Huang et al. 2015).

However, the role of physical and biogeochemical processes in shaping microbial communities in the oligotrophic ocean remains unclear. While some general *in situ* trends are

apparent— for example the predominance of *Synechococcus* and PPE in nutrient rich waters and the predominance of *Prochlorococcus* in nutrient depleted regions—spatial and temporal variability provide significant challenges to the generalization of these patterns at a global level (Fuhrman 2009). And while there are accounts of microbial community structure in oligotrophic regions of the North Atlantic (Partensky et al. 1996) and North Pacific (Campbell and Vaulot 1993; Karl 1999), as well as for the Mediterranean (Denis et al. 2010) and Arabian Seas (Campbell et al. 1998), the South Pacific remains less well documented. Although there are three reports on the distribution of bacteria, *Prochlorococcus*, *Synechococcus* and PPE abundance and biomass in the oligotrophic subtropical southeast Pacific (Grob et al. 2007a; b; Lepère et al. 2009), and two papers describing cell abundance distribution of *Prochlorococcus*, *Synechococcus* and PPE in the southwest Pacific near New Caledonia (Blanchot and Rodier 1996; Barboza Tenório et al. 2018), there are no reports on community structure in the oligotrophic regions of the Western Tropical South Pacific (WTSP). Moreover, despite their crucial role as grazers of microbial plankton, small-sized heterotrophic protists (e.g., non-pigmented cells between ~2 to 5 μm, hereafter HNF), also called flagellates, have received little attention (Christaki et al. 2011). Although there have been an increasing number of reports focusing on their role as predators over the last couple of decades, HNF are not routinely measured and their distribution relative to other microbial groups is not well constrained (Christaki et al. 2011).

In this study, we present an account of microbial plankton community structure during late austral summer in distinct biogeochemical regions of the WTSP, ranging from mesotrophy to ultraoligotrophy. Our primary goal is to document the microbial community structure in the region, based on flow cytometry data capturing bacterioplankton (low-DNA-content, LNA, and

high-DNA-content, HNA), phytoplankton (*Prochlorococcus*, *Synechococcus* and PPE) and HNF groups. In addition, we describe the dominant biogeochemical gradients observed along the cruise transect, and attempt to identify the physical and ecological variables influencing the abundance and distribution of plankton groups in the region. We also compare several previously published empirical models that make use of bacteria and HNF abundances to evaluate trophic interactions between populations of heterotrophic plankton groups.

4.3. Materials and Methods

4.3.1. Field sampling

A zonal characterization of the biogeochemistry and biological diversity of the western tropical South Pacific (WTSP) was conducted along trophic gradients during the OUTPACE cruise (Oligotrophy to UlTra-oligotrophy PACific Experiment, DOI: http://dx.doi.org/10.17600/15000900, RV *L'Atalante*, February–April 2015) between New Caledonia and Tahiti (Moutin et al. 2017). To describe the longitudinal and vertical distribution of different groups of pico- and nanoplankton, we sampled 15 short duration stations (SD1 to SD15, occupied during 8 h, Figure 4.1). Three long duration stations (LDA, LDB, and LDC, occupied during 7 days), chosen for their contrasted biogeochemical conditions (Table 4.1), were sampled in Lagrangian mode (de Verneil et al. 2018). In keeping with Moutin et al. (2017), stations sampling the waters of the Melanesian Archipelago (SD1 to SD12 and LDA) were classified as MA, while stations sampling the western South Pacific Gyre (SD13 to SD15 and LDC) were classified as GY. Long duration station B was in the late stages of a phytoplankton bloom at the time of sampling (de Verneil et al. 2017). As such, the station was analyzed independently of other stations on the transect and is simply referred to as LDB. All stations were sampled at 12 different depths from the surface down to 200 m for microbial characterization of the extended photic layer. The photic layer, Z_e (m) is defined as the sunlit

layer of the water column between the surface and the depth where irradiance is reduced to 0.1 % of its surface value (hereafter $PAR_{0.1}$).

4.3.2. Pico- and nano-plankton analyses

For cell enumeration, duplicate 1.8-ml samples were fixed (0.25 % electron microscopy grade paraformaldehyde, w/v) for 10-15 min at room temperature and in the dark, flash-frozen in liquid nitrogen and stored at -80 °C for later analysis using a BD Influx flow cytometer (BD Biosciences, San Jose, CA, USA). Pigmented groups, *Prochlorococcus*, *Synechococcus* and PPE, were enumerated in unstained samples for 5 min at ~61 µl/min. Bacteria were discriminated in a sample aliquot stained with SYBR Green I DNA dye (1:10,000 final) and enumerated for 1 min at ~65 µl/min. HNF were analyzed for 8 min at ~193 µl/min in a sample stained with SYBR Green I at 1:5,000 final concentration (Zubkov et al. 2007; Christaki et al. 2011). Particles were excited at 488 nm (plus 457 nm for unstained samples). Forward (< 15°) scatter (FSC), side (90°) scatter (SSC), green fluorescence (530/40 nm), orange fluorescence (580/30 nm) and red fluorescence (> 650 nm) emissions were measured. Pigmented groups were identified and enumerated based on their chlorophyll (red) fluorescence and FSC (size) signatures. The high phycoerythrin (orange) signal in *Synechococcus* was used to distinguish them from *Prochlorococcus* and PPE. Using a FSC detector with small particle option and focusing a 488 plus a 457 nm (200 and 300 mW solid state, respectively) laser into the same pin-hole greatly improved the resolution of dim surface *Prochlorococcus* population from background noise (Duhamel et al. 2014). LNA and HNA bacteria were discriminated based on their low and high green fluorescence, respectively, in a SSC vs green fluorescence plot (Vazquez-Dominguez et al. 1999; Van Wambeke et al. 2011). In samples from the upper euphotic layer, where *Prochlorococcus* signal overlapped with HNA in an SSC or FSC vs green fluorescence plot, *Prochlorococcus* abundance counted in unstained samples were subtracted

from the HNA abundance enumerated in a larger gate. Total bacteria refers to the sum of LNA and HNA abundances. Reference beads (Fluoresbrite, YG, 1 μm) were added to each sample and red fluorescence from chlorophyll and FSC values are presented relative to the reference beads (arbitrary units, A.U.).

N_2 fixation rates were measured in triplicate at all stations (except SD13) using the $^{15}N_2$ isotopic tracer technique (adapted from Montoya et al., 1996). Briefly, seawater samples were collected in HCl-washed, sample-rinsed (3 times) light- transparent polycarbonate 2.3-L bottles from 6 depths (75 %, 50 %, 20 %, 10 %, 1 %, and 0.1% surface irradiance levels), sealed with caps fitted with silicon septa and amended with 2 mL of 98.9 atom % $^{15}N_2$ (Cambridge isotopes). Incubation bottles were incubated in on-deck incubators equipped with circulating seawater at the specified irradiances using blue screening. Incubations were stopped by filtration of the entire sample onto precombusted 25-mm GF/F glass fiber (Whatman, 0.7 μm nominal pore size) filters, which were then analyzed for $^{15}N/^{14}N$ ratios and PON concentrations using an elemental analyzer coupled to a mass spectrometer (EA-IRMS, Integra CN, SerCon Ltd) as described in Bonnet et al. (2018).

4.3.3. Data analysis and statistics

All data analyses were performed using R Studio (R Core Team, 2016). All figures were prepared using the ggplot2 package (Wickham 2009), with the exception of the contour plots presented in Figure 4.2, which were prepared using Ocean Data View 4.7.10 (Schlitzer 2017).

Chlorophyll fluorescence of microbial groups was calculated as the ratio of mean red fluorescence of each cell population to that of the reference beads, based on flow cytometry results. Per-cell biomass was calculated using previously published conversion factors: 29 fg C per *Prochlorococcus* cell, 100 fg C per *Synechococcus* cell, 11.5 fg C per bacterial cell, and 1500 fg C per PPE cell (Zubkov et al. 2000). Biomass contributions of different phytoplankton groups

were estimated by multiplying cell abundances by the conversion factors above and integrating by depth across the euphotic zone. To account for larger eukaryotes not captured by cytometry, we used the method described in Vidussi et al. (2001) to estimate the relative biomass contributions of diatoms and dinoflagellates based on concentrations of fucoxanthin and peridinin relative to those of total diagnostic pigments: zeaxanthin, DV-*Chla*, Chlorophyll *b* (Chl*b*), 19' hexanoyloxyfucoxanthin (19'-HF), 19' butanoyloxyfucoxanthin (19'-BF), alloxanthin, fucoxanthin, and peridinin. Pigment concentrations were measured via high performance liquid chromatography, as described elsewhere (Ras et al. 2008).

Concentrations of PO_4, NO_2, and NO_3 were measured using a SEAL AA3 HR auto-analyzer (SEAL Analytical, UK), as described by Moutin et al. (2018). Abundances, biomass, and nutrient values are reported as either depth-integrated totals or as depth-normalized averages calculated by dividing depth-integrated totals by the depth of integration. To account for large vertical gradients in both abundances and nutrient concentrations within the Ze, integrations were performed across two depth ranges: the upper photic zone (Z_u), integrating from the surface to the recorded 2.7 % isolume (hereafter $PAR_{2.7}$) and the lower euphotic zone (Z_l), integrating from $PAR_{2.7}$ to the recorded 0.1 % isolume (hereafter $PAR_{0.1}$). The mixed layer depth was measured continuously along the transect, as described by de Verneil, et al. (2017). The mixed layer was entirely within Z_u at all stations (Figure 4.2), while the nitracline, defined as the depth where measurements of $NO_2 + NO_3$ (hereafter NO_x) first exceed 0.1 µM, occurred within Z_l at all stations (Figure 4.2F). Attenuation coefficients, k, were calculated by using CTD PAR measurements to solve the Beer-Lambert equation between surface PAR and that corresponding to $PAR_{0.1}$.

To maximize the power of statistical tests, depth-integrated values were calculated for

individual casts, without averaging results from long duration stations. Because each cast was made on a different day of occupation, doing so did not violate assumptions of independence during subsequent statistical analyses. Integration results were then assigned categorical variables corresponding to biogeochemical condition or euphotic zone, with two-way ANOVA being used to identify statistical relationships between parameters according to these variables (Table 4.2). Tukey's Honest Significant Difference post-hoc test was used to compare group means when two-way ANOVA indicated significant between-group differences. To ensure that data met the requirements for ANOVA (e.g., normally distributed, and with equal variance between groups), values were log transformed before performing statistical tests. In the case of nutrient data and N_2 fixation rates, which were often skewed by large occurrences of small values, data were transformed using the formula data = Log (data*100 + 1). Bivariate comparisons between biogeochemical parameters were performed using Pearson's correlation. The Shapiro-Wilk test was used to assess normality, while the Levene's test was used to confirm homogeneity of variance.

4.4. Results

4.4.1. Biogeochemical characteristics of study area

Stations along the transect were characterized by warm sea surface temperatures (mean 29.4 ± 0.4 °C). The water column was strongly stratified along the entire transect, with mixed layer depths ranging from 21 ± 5 m for MA to 25 ± 8 m for GY (Figure 4.2). Conversely, there was a significant west to east decrease in light attenuation (k), ranging from 0.059 ± 0.006 m^{-1} for MA to 0.044 ± 0.005 m^{-1} for GY. An exception to this trend was found at LDB, where k increased to 0.078 ± 0.021. Corresponding to these changes in k, $PAR_{0.1}$ deepened from west to east, ranging from 113 ± 13 m for MA to 178 ± 5 m for GY. Again, LDB presented an exception to this general trend, where $PAR_{0.1}$ was recorded at 83 ± 5 m.

All stations across the transect featured a prominent deep chlorophyll maximum (DCM). Mirroring changes in *k*, the DCM showed a general increase in depth from west to east, ranging from 85 ± 20 m to 133 ± 20 m from MA to GY, respectively. The DCM depth at LDB was also an exception, decreasing to 50 ± 19 m. The concentration of dissolved oxygen was near equilibrium with the atmosphere near the surface, becoming slightly oversaturated below the mixed layer. This subsurface maximum occurred at a mean depth of 55 ± 18 m and was weakly correlated with the depth of the DCM (Pearson's $r = 0.44$, $p < 0.01$). Oxygen levels within Z_e were above 158.31 mol kg^{-1} across the entire transect, with there being no suboxic regions at any of the stations sampled. The nitracline generally tracked the DCM, occurring at depths ranging from 93 ± 17 m for MA to 127 ± 13 m for GY. The nitracline was decoupled from the DCM at LDB, where it occurred at 108 ± 22 m.

NO_X concentrations were depleted in Z_u across all biogeochemical conditions. While depth-normalized NO_X concentrations were significantly elevated in Z_l for MA and GY, (ANOVA, $p < 0.01$), no significant difference was encountered between Z_u and Z_l at LDB (Table 4.1). Depth-normalized phosphate (PO_4) concentrations in Z_u, were significantly elevated at GY relative to other stations (ANOVA, $p < 0.01$), although no significant differences were identified between MA and LDB.

4.4.2. Phytoplankton community structure

Prochlorococcus dominated phytoplankton abundances at all sampling sites, with average Z_e-integrated abundances being two orders of magnitude greater than those of *Synechococcus* and PPE (Table 4.1). Z_e-integrated *Prochlorococcus* abundances ranged from 135 $x10^{11}$ cells m^{-2} at SD3 to 283 $x10^{11}$ cells m^{-2} at SD1, while those for *Synechococcus* ranged from 0.65 $x10^{11}$ cells m^{-2} at SD15 to 18.62 $x10^{11}$ cells m^{-2} at LDB. PPE abundances ranged from 1.40 $x10^{11}$ to 2.60 $x10^{11}$ cells m^{-2} at SD3 and SD12, respectively. There were no significant differences in Z_e-

integrated abundances of these groups across biogeochemical conditions, except for those of *Synechococcus*, which were significantly greater at LDB compared to MA or GY (ANOVA, $p < 0.01$). Transect-wide, *Prochlorococcus* accounted for approximately 97 ± 2 % of total phytoplankton cells enumerated by flow cytometry. *Synechococcus* and PPE accounted for 2 ± 2 % and 0.8 ± 0.2 % of total phytoplankton abundance, respectively. Pooling all data, statistically significant correlations were identified between all pairs of plankton groups (Pearson's R, $p < 0.01$).

Based on Z_e-integrated abundances, relative contributions of different phytoplankton groups to total phytoplankton abundance showed considerable longitudinal variation. *Synechococcus* accounted for 0.4 ± 0.1 % of phytoplankton cells at GY, 2.4 ± 1.9 % at MA, and 7.4 ± 4.4 % at LDB. P*rochlorococcus* abundances, by contrast, represented 92.1 ± 4.3 % of phytoplankton cells at LDB and, 96.7 ± 2.0 % at MA, and 98.6 ± 0.2 % at GY. Ratios of Z_e-integrated abundances of *Prochlorococcus* to *Synechococcus* varied significantly (ANOVA, $p < 0.01$) between GY (235.7 ± 65.1) and LDB (16.2 ± 10). PPE abundances showed less variability, with relative abundances ranging from 0.4 ± 0.1 % to 0.9 ± 0.1 % of phytoplankton cells at LDB and GY, respectively. Differences in the relative abundance of PPE between biogeochemical conditions were not statistically significant. Statistically significant negative correlations were found between concentrations of NO_X and PO_4 and all plankton groups (Pearson's R, $p < 0.01$), while significant positive correlations were identified between N_2 fixation rates and abundances of *Prochlorococcus* and heterotrophic bacteria (Pearson's R, $p < 0.01$). These correlations persisted when subsetting data to include mixed layer values alone (Pearson's R, $p < 0.01$), except for correlations between NO_x and plankton groups.

Most stations were characterized by a well-defined two-tier distribution of phytoplankton

within the Z_e (Figure 4.3), with depth-integrated abundances of *Prochlorococcus* and *Synechococcus* being greatest in the Z_u, and PPE abundances being greatest in Z_l (Table 4.1). These differences between Z_u and Z_l abundances were found to be statistically significant for *Prochlorococcus* across all conditions. Differences were significant for *Synechococcus* at MA, and for PPE at MA and GY (ANOVA, $p < 0.01$). *Prochlorococcus* and PPE abundances showed subsurface maxima at both MA and GY. Averaging across the transect, *Prochlorococcus* abundance maxima occurred at depths corresponding to 24.2 ± 24.4 % PAR while PPE maxima occurred at depths corresponding to 0.6 ± 0.4 % PAR. Depths of these maxima showed a west to east increase, and were significantly deeper at GY than at MA for all phytoplankton groups (ANOVA, $p < 0.01$).

There was no significant variation in Z_e integrated biomass between different conditions, although Z_u-integrated biomass was significantly greater ($p < 0.01$) at LDB compared to MA and GY stations (Figure 4.4). In keeping with relative abundances, *Prochlorococcus* cells represented the greatest fraction of Z_e biomass, accounting for an average of 77.1 ± 5.5 % across the transect. By comparison, PPE accounted for an average of 18.7 ± 5.4 % of Z_e biomass, while *Synechococcus* accounted for 3.9 ± 4.3 %. However, there was considerable vertical and longitudinal variation in these trends, especially in contributions to total biomass by PPE and *Synechococcus* populations (Figure 4.4). PPE accounted for 29 ± 14 % of phytoplankton biomass considering Z_l alone, and up to 64 % of Z_l biomass at SD4. *Synechococcus* accounted for up to 13 % of Z_u phytoplankton biomass at LDB.

4.4.3. Distributions of bacterioplankton and HNF

Z_e-integrated bacterial abundances ranged from 4.17 x 10^{11} to 6.61 x10^{11} cells m^{-2} at LDA and LDB, respectively (Table 4.1). Despite this range, there was relatively little variation when

comparing biogeochemical regions; while average Z_e-integrated abundances at GY were somewhat elevated compared to those for MA; and while those at LDB were amongst highest on the transect, these differences were not statistically significant. Examining HNA and LNA subpopulations, Z_e integrated abundances for HNA bacteria ranged from 115 x 10^{11} to 291 x 10^{11} cells m^{-2} at SD3 and SD9, respectively, while values for LNA ranged from 155 x10^{11} cells m^{-2} at SD3 to 298 x10^{11} cells m^{-2} at SD8. As with total bacteria, there were no statistically significant longitudinal differences in Z_e integrated HNA or LNA abundances when comparing different biogeochemical regions. The fraction of HNA to total bacteria (%HNA) ranged from 41.1 ± 2.1 % to 48.0 ± 4.9 % between GY and LDB, respectively. Values for %HNA were significantly greater at LDB relative to MA and GY (ANOVA, $p < 0.01$) and at MA relative to GY (ANOVA, $p < 0.01$). Bacterial abundances showed less variability with depth than did phytoplankton groups, with there being no significant differences in depth-integrated abundances between Z_u and Z_l for total bacteria, HNA, or LNA. Depth profiles of %HNA, however, were more variable than those for total bacteria. %HNA increased from the surface to $PAR_{0.1}$ across all biogeochemical regions, and distinct local minima were apparent near the DCM at MA and GY (Figure 4.5).

Mean HNF abundances in Z_u ranged from 0.38 x 10^3 to 2.3 x 10^3 cells ml^{-1} at SD 15 and LDB, respectively. Z_e integrated abundances at LDB were significantly greater than those at GY (ANOVA, $p < 0.01$), although no significant differences were found between MA, GY or LDB. Depth integrated abundances of HNF were significantly greater in Z_u than Z_l at MA and GY (ANOVA, $p < 0.01$), while there was no significant change in HNF abundances with depth at LDB.

4.4.4. Bottom-up vs. top-down control of microbial communities

To assess the roles of top-down and bottom-up control over microbial group abundances along the transect, we used a combination of approaches based on previously published models. The model described by Gasol (1994) was used to assess top-down vs bottom-up control of HNF abundance (Figure 4.6). Specifically, this approach compares observed ratios of bacteria to HNF with HNF abundance maxima estimated from empirical data and theoretical interactions between bacteria and HNF. The main assumption of the model is that bacteria to HNF ratios nearer to theoretical maxima implies increased bottom-up control of HNF by bacterial abundance. This difference is quantified with the parameter *d*, which is calculated as the difference between theoretical and observed HNF abundances. Small values of *d* are thus interpreted as being indicative of top-down control on bacterial populations by HNF, or by a significant use of resource other than bacteria by HNF. Large values of *d* are interpreted as being indicative of a decoupling between the two groups, and/or a top-down control of HNF by their predators, like ciliates. Average Z_u values for *d* were 0.59 ± 0.11, 0.62 ± 0.19, and 0.80 ± 0.09 for MA, LDB, and GY, respectively. By contrast, average Z_l values for *d* were 0.46 ± 0.14, 0.43 ± 0.08, and 0.45 ± 0.10 for MA, LDB, and GY, respectively. Z_u Values for *d* were significantly elevated at GY relative to MA and LDB (ANOVA, $p < 0.01$). No significant differences in *d* were identified between biogeochemical regions in Z_l.

Regressions between abundances of bacteria and HNF were measured for Z_u and Z_l across biogeochemical conditions (data not shown). The variability in HNF abundances explained by bacteria abundance was generally greater in Z_l compared to Z_u. Z_u bacterial abundances explained 24 %, 30 % and 30 % of variability in HNF abundance at MA, LDB, and GY, respectively. In Z_l, bacterial abundances explained 57 % and 72 % of variability at MA and GY stations, respectively, while this relationship was not statistically significant at LDB. Repeating this

procedure for HNA bacteria alone, Z_u HNA abundances were found to explain 15 %, 29 %, and 73 % of variability in HNF abundance at MA, LDB, and GY, respectively. Z_l bacteria abundances were found to explain 61 % of variability in HNF populations at GY, while relationships at MA and LDB were insignificant. Z_u values for %HNA explained 73 % of variability in HNF abundances at GY, while this relationship was weak and insignificant at MA and LDB.

Using the ciliate abundances collected by Dolan et al. (2016) during the OUTPACE cruise, ratios of depth-integrated abundances of ciliates to HNF (with HNF abundances multiplied by 10^{-11} for readability) were found to range from 2.8 in the upper euphotic zone at LDB to 17.6 in the lower euphotic zone at LDC (Figure 4.2E). In the upper euphotic zone, this ratio increased from 2.9 at LDB to 10.0 at LDC and 10.9 at LDA. The lower euphotic zone showed a slightly different pattern, with the ratio increasing from 9.0 at LDA to 9.6 at LDB and 17.5 at LDC. Because data available were limited to one set of measurements at each of those three stations, it was not possible to determine whether differences in these results were statistically significant. However, comparing differences between biogeochemical conditions based on observations at individual depths, rather than depth-integrated values, indicated the ratio of ciliates to HNF to be significantly lower at LDB compared to LDC (ANOVA, $p < 0.01$). Differences between other biogeochemical conditions, however, were statistically insignificant. No significant vertical or longitudinal differences were identified for ratios of HNF to cyanobacteria, nor for ratios of bacteria to cyanobacteria.

4.4.5. Distribution of pigments and photo acclimation in different phytoplankton groups

Phytoplankton group specific relative fluorescence values obtained by flow cytometry for *Prochlorococcus*, *Synechococcus*, and PPE, showed significant (t-test, $p < 0.01$) increases with depth across all biogeochemical conditions (Figure 4.3). Phytoplankton relative fluorescence for

all groups showed little variation within the Z_u, although marked increases occurred in the region of $PAR_{2.7}$. *Prochlorococcus* relative fluorescence showed a continuous increase to 200 m at MA and GY, and an increase to 150 m at LDB. *Synechococcus* and PPE showed clear maxima near or just below $PAR_{0.1}$, although PPE relative fluorescence showed a continuous increase to 200 m at GY.

Analysis of HPLC pigments data using the approach described in Vidussi et al. (2001) largely mirrored our flow cytometry results. Transect-wide, zeaxanthin and chlorophyll *b*, pigments corresponding to cyanobacteria and prochlorophytes dominated in Z_u, accounted for 80 ± 5.1 % of total diagnostic pigments. Fucoxanthin and peridinin, diagnostic of diatoms and dinoflagellates, accounted for 3.8 ± 1.0 %. Concentrations of 19'HF and 19'BF—diagnostic pigments typically used to assess abundances of prymnesiophytes and chrysophytes/pelagophytes, respectively (Wright and Jeffrey, 2006)—showed significant horizontal and vertical variability (Table 4.3). Absolute concentrations of both pigments showed significant increases with depth at MA and GY (ANOVA, $p < 0.01$), although increases at LDB were not statistically significant. Ratios of 19'HF:Chl*a* were significantly greater than those of 19'BF:Chl*a* across all biogeochemical conditions (t-test, $p < 0.01$). Ratios of 19'HF:Chl*a* showed significant increases with depth at MA (ANOVA, $p < 0.01$), although increases at GY and LDB were not statistically significant. Ratios of 19'BF:TChl*a* showed significant increases at MA and GY (ANOVA, $p < 0.01$), while increases at LDB were insignificant. Z_l ratios of 19'HF:19'BF were significantly elevated at LDB compared to MA (ANOVA, $p < 0.01$). That no similar such difference was observed in comparing LDB to GY is likely the result of the reduced number of samples available for making this comparison. Indeed, the difference was nearly significant (ANOVA, $p = 0.03$), while Z_l ratios of 19'HF:19'BF at MA were remarkably similar (ANOVA, p

= 0.99). A moderately strong relationship was identified between carotenoid concentrations and PPE abundances, with variability in PPE accounting for 46 % of variability in 19'BF + 19'HF ($p < 0.01$).

4.5. Discussion

4.5.1. Distribution of phytoplankton populations in the WTSP and relative contribution to biomass

Transect-wide biogeochemical conditions captured by our data were similar to those of the "typical tropical structure" described by Herbland and Voituriez (1979), featuring large abundances of pico-sized organisms in Z_u, a deep nitracline, and a prominent DCM in Z_l. Differences in relative abundances of phytoplankton groups between Z_u and Z_l showed a clear two-tiered vertical niche partition, with *Prochlorococcus* and *Synechococcus* reaching maximum abundances in the Z_u and PPE achieving maxi- mum abundances in the Z_l. This vertical distribution has been well documented in other regions and is thought to be characteristic of highly stratified oligotrophic systems (Partensky et al. 1996; Dore et al. 2008; Painter et al. 2014). Based on estimates from HPLC data, larger organisms such as diatoms and dinoflagellates, were present in very low abundance along the transect, in comparison to small-sized phytoplankton.

Although the use of different conversion factors for estimating per-cell carbon makes it difficult to compare between different studies, our biomass estimates largely agree with those reported for other oligotrophic regions (Partensky et al. 1996; Zubkov et al. 2000; Pérez et al. 2006; Grob et al. 2007b). *Prochlorococcus* accounted for the large majority of phytoplankton biomass in the Z_u, with *Synechococcus* and PPE only making relatively minor contributions. In Z_l, by contrast, PPE accounted for a more sizeable and occasionally dominant share of phytoplankton biomass. This effect was particularly pronounced at GY, where increases in Z_l

PPE biomass compensated for reductions in Z_u *Prochlorococcus* biomass, resulting in Z_e biomass totals to be similar to those at MA.

To compare our results to those from other ocean basins, we conducted a meta-analysis of datasets reporting *Prochlorococcus*, *Synechococcus*, and PPE abundances alongside Chl*a* concentrations. Mean depth-integrated abundances of *Synechococcus* and PPE measured along the OUTPACE transect were similar to those reported elsewhere, as were those of *Prochlorococcus* at MA and GY (Table 4.4). However, mean depth-integrated abundances of *Prochlorococcus* at LDB were considerably higher than mean values for other regions. Some of this variation may result from the instrumentation used, with earlier cytometers generally being thought to underestimate weakly fluorescent *Prochlorococcus* cells near the surface. Regardless, with Z_e integrated *Prochlorococcus* abundances at LDB being greater than any others encountered in the literature, these results highlight the importance of transient, localized blooms to cyanobacterial abundance in the region.

The observed increase in the percentage of Z_e-integrated *Prochlorococcus* abundances accounting for total picophytoplankton abundance from LDB to GY captures a global trend. The proportion of *Prochlorococcus* cells accounting for total picophytoplankton abundances have generally been reported to decrease with increased Chl*a* concentrations, while those of *Synechococcus* cells have been found to increase along the same gradient (Table 4.4). In the Sargasso Sea, where winter mixing allows for the resupply of surface nutrients, long-term studies have captured this relationship as a seasonal pattern, with ratios of *Prochlorococcus* to *Synechococcus* increasing inversely with changes in the depth of the nitracline (Campbell et al. 1998; Durand et al. 2001). The same phenomenon has been reported along biogeochemical conditions in the North Atlantic (Partensky et al. 1996; Zubkov et al. 2000), as well as the South

East Pacific (Grob et al. 2007a; Rii et al. 2016b). By contrast, sites with comparatively limited seasonal variability, like Station ALOHA in the North Pacific subtropical gyre, have shown consistently high ratios of *Prochlorococcus* to *Synechococcus* ratios year-round (Rii et al. 2016a), while studies of nitrate-rich eutrophic regions often report the complete exclusion of *Prochlorococcus* by *Synechococcus* and eukaryotic populations (Sherr et al. 2005; Zubkov et al. 2007). The high relative abundance of *Prochlorococcus* at LDB compared to sites with similar Chl*a* concentrations may be due to having captured the bloom in decay (de Verneil et al. 2017), with nutrients having been largely depleted at the time of measurement and relative *Prochlorococcus* abundances returning to levels more representative of the WTSP. Alternatively, the bloom conditions may be incomparable to regions with more persistent inputs of nutrients.

4.5.2. Factors regulating the horizontal distribution of phytoplankton groups

In examining the potential factors regulating the distribution and abundance of cyanobacterial groups, our data did not reward any expectations that abundances of *Prochlorococcus*, *Synechococcus* or PPE would correlate meaningfully with NO_x concentrations. While negative relationships were identified between plankton abundances and NO_x concentrations across all sites, this was likely the result of changes in these parameters with depth rather than being indicative of any causal relationship. Indeed, comparing values in Z_u alone, correlations between NO_x and plankton abundances became insignificant, with NO_x being largely depleted above $PAR_{2.7}$. However, correlations between N_2 fixation, PO_4 concentrations, and plankton abundances persisted even when subsetting data to only include measurements within the mixed layer, indicating covariation between these parameters to occur independently of depth. Specifically, correlations between N_2 fixation rates and abundances of cyanobacteria suggest plankton abundances in the surface to respond to diazotroph-derived nitrogen (ammonia and DON) provided by N_2-fixing organisms, notably *Trichodesmium* which dominated in the upper euphotic

zone at MA (Stenegren et al. 2017). Previous studies have demonstrated growth to increase with DON enrichment in both *Synechococcus* and *Prochlorococcus* cultures (Moore et al., 2002), while others have indicated that diazotrophs may provide a large enough input of fixed nitrogen to sustain large populations of cyanobacteria (Bonnet et al. 2016a). Moreover, previous experiments in the New Caledonia lagoon have shown a rapid transfer (24-48 h) of recently fixed N by *Trichodesmium* towards non diazotrophic phytoplankton and heterotrophic bacteria (Bonnet et al. 2016b), while Caffin et al. have demonstrated the efficient transfer of N fixed by UCYN-B bacteria to the planktonic food web along the OUTPACE transect (Caffin et al. 2018). Biological nitrogen inputs may allow for a more complete utilization of PO_4 at sites featuring high nitrogen-fixation rates (Mather et al. 2008; Moutin et al. 2018), accounting for the negative correlations observed between PO_4 concentrations and abundances of *Prochlorococcus* and *Synechococcus,* as well as for the negative correlations observed between PO_4 concentrations and N_2 fixation rates. These results, along with the low DIP turnover rates reported, suggest intense competition for phosphorus within the mixed layer, and a rapid transfer of fixed N toward heterotrophic bacteria (Van Wambeke et al. 2018).

4.5.3. Factors regulating vertical variability in phytoplankton community structure

Considerable variation in vertical distributions of phytoplankton groups was observed between biogeochemical regions. Although *Synechococcus* and PPE appeared confined to high-light and low-light depths, respectively, *Prochlorococcus* abundances showed a greater deal of vertical variability, with *Prochlorococcus* subsurface abundance maxima varying widely with respect to PAR (Figure 4.3). This indicates *Prochlorococcus* distributions in Z_u to be less sensitive to changes in light availability than other phytoplankton groups (Partensky et al. 1999), possibly as a result of comparatively reduced increases in per-cell chlorophyll concentrations with depth, as reflected by relative fluorescence values (Figure 4.3). However, the observed *Prochlorococcus*

distributions reflects the average distribution of a mosaic of different ecotypes encompassing high diversity in their response to nutrients, light, and temperature (Moore et al. 2002; Johnson 2006; Kashtan 014). Indeed, the increase in depth of *Prochlorococcus* abundance maxima observed at GY is likely the result of the deepening of the euphotic layer, combined with the reduction of high-light ecotypes in Z_u. While previous studies have reported correlations between *Prochlorococcus* abundance maxima and nitracline depth (Olson et al. 1990; Li 1995), no similar such patterns were observed in our data. These distributions may be a transient feature formed during restratification following winter mixing (Partensky et al. 1999), and are unlikely to be in response to nitrate availability, given the small nitrate utilization by *Prochlorococcus* in natural samples (Casey et al. 2007). These results may also be the consequence of the difficulty in detecting weakly fluorescent high-light *Prochlorococcus* following earlier flow cytometry protocols with less sensitive instruments.

In contrast to other phytoplankton groups, PPE abundances were marginal in Z_u, but increased dramatically below PAR_{10}, reaching maximal abundances at depths closely correlated with those of the DCM. The lack of variability of PPE abundance maxima relative to PAR, along with the decoupling of PPE maxima from the nitracline at LDB, suggest PPE abundances to be primarily controlled by light levels rather than by the availability of dissolved nutrients. However, it is difficult to consider these factors independently with the increased chlorophyll concentrations required at low light levels likely increasing nitrogen requirements on shade-adapted organisms (Edwards et al. 2015).

Differences in vertical distributions of PPE between stations likely also reflects variability in the composition of PPE communities. The decrease with depth observed in ratios of 19'HF:19'BF would suggest prymnesiophytes to dominate in the Z_u, with chrysophytes and pelagophytes

accounting for a greater proportion of total PPE abundance in Z_l. Similar distributions have been reported elsewhere and have been suggested to reflect control of chrysophyte and pelagophyte abundances by nitrate avail- ability (Barlow et al., 1997; Claustre et al., 1994; Marty et al., 2002). This interpretation agrees with our results, where the separation of PPE abundance maxima from the nitracline coincided with significantly elevated Z_l ratios of 19'HF:19'BF compared to MA. That elevated Z_l values for 19'HF:19'BF at LDB coincide with transect-wide maximal concentrations of NH_4 suggest that prymnesiophytes may preferentially utilize reduced forms of nitrogen. This would also account for the elevated abundances of this group in Z_u across all biogeochemical conditions, where reduced forms of nitrogen would generally be expected to be more abundant as a result of nutrient recycling. Admittedly, with variability in 19'HF+19'BF only accounting for ~50 percent of variability in picoeukaryote abundances, it is likely that observed patterns of 19'HF:19'BF capture changes in nano- and micro-sized eukaryotes in addition to PPE.

It is also possible that Z_l PPE populations are responding to the availability of nitrogen fixed by UCYN-A cyanobacteria, which were reported to have distributions at least qualitatively similar to those of PPE across the transect (Stenegren et al. 2017). Several UCYN-A clades have been identified to form symbioses with small prymnesiophytes, including at least one pico-sized haptophyte (Martínez-Pérez et al. 2016), making it seem plausible that such relationships could play an important role in controlling PPE distributions.

4.5.4. Factors controlling bacterial abundance and the role of HNF

Average bacterioplankton abundances in the Z_e ($3.6 \times 10^8 \pm 2.6 \times 10^5$ cells l^{-1}) were within the established range of 1-5 x 10^8 cells l^{-1} for the oligotrophic ocean (Ducklow 2002). That bacterioplankton abundances at LDB should exceed this range, slightly exceeding 7.0 x 10^8 cells l^{-1}, is not surprising, based on the increased abundances of phytoplankton at that station. Surface

values for %HNA, ranging from 30 % at SD15 to 51 % at SD1, are similar to previously reported values, as is the observed increase in %HNA with depth (Van Wambeke et al. 2011). The reduced values for %HNA reported for GY are similar to those reported for other nutrient-limited regions and may be the result of LNA cells being capable of more rapid growth than HNA under nutrient limitation (Nishimura et al. 2005; Andrade et al. 2007).

To assess variation in trophic interactions between HNF and bacteria across biogeochemical conditions, we used the method presented by Gasol (1994), which compares observed ratios of HNF to bacteria with theoretical maxima (*d*). Previous applications of this model have demonstrated an increase in top-down control of bacterial populations in low-chlorophyll regions, demonstrated by low values for *d* under nutrient limitation (Gasol et al. 2002). These results find general support in the literature, on the underlying assumption that nutrient-limited regions are characterized by reduced abundances of top predators, resulting in increased grazing pressure on bacteria via a trophic cascade (Pernthaler 2005). Our data, however, contrast with these conclusions, with values for *d* being significantly greater at GY relative to those corresponding to MA or LDB, suggesting a reduction in grazing pressure on heterotrophic bacteria with increased nutrient limitation in Z_u.

There are several possible explanations for this result. Based on the significantly reduced Z_u values for %HNA encountered at GY, decoupling of bacterial and HNF populations may be the result of diminished prey quality at these sites. While it has been debated whether HNA and LNA can be interpreted as representing active and inactive cells, respectively (Vaqué et al. 2001; Jochem et al. 2004), HNA bacteria have generally been found to be larger in diameter than LNA bacteria (Van Wambeke et al. 2011), possibly making them more susceptible to grazing. Such a phenomenon has been described previously in nutrient- limited regions (Vaqué et al. 2001;

Longnecker et al. 2010); although studies conducted in relatively nutrient-rich regions have reported no similar such evidence (Jochem et al. 2004). In our data, the increased R^2 values for linear regressions between abundances of HNA and HNF at GY suggest that HNA densities may be more important in controlling HNF densities at GY than at MA or LDB. This may account for the increased values of *d* calculated at GY—the relationships used to establish the model's theoretical maxima not accounting for changes in grazing rates in response to %HNA, and thereby potentially overestimating HNF abundance, as derived from bacterial abundance, in cases where %HNA is low. Plotting bacterial abundances against the bacterial production data reported by Van Wambeke *et al.* (2018), and interpreting regression slopes using the criteria described by Ducklow *et al.* (1992) in Z_u, we found no evidence for bottom-up control of bacteria populations at MA, LDB or GY (data not shown). This is similar to the results obtained by Morán *et al*. (2017), who applied the Ducklow model to data collected in the South Pacific during austral summer and reported very weak bottom-up control at all sampling sites, calculating regression slopes around 0.2 for samples between the surface and 4000 m. The authors likely would have obtained still lower slopes had their analysis been restricted to surface data alone, as we found for Z_u.

The increase in *d* at GY could alternatively result from increased grazing on cyanobacterial prey, given that the HNF abundances predicted by the Gasol model are calculated on the assumption that HNF only graze on heterotrophic bacteria. Previous studies have reported HNF to graze on cyanobacteria, generally at rates similar to those reported for grazing on heterotrophic bacteria (Ferrier-Pagès and Gattuso 1998; Christaki 2001; Cuevas and Morales 2006). However, ratios of bacteria to cyanobacteria are largely invariable across the transect, as are ratios of HNF to cyanobacteria. Both of these values would reasonably be expected to vary if

responsible for the reported differences in *d*. To account for the possibility that *Prochlorococcus* cells were erroneously identified as heterotrophic bacteria at the time of the model's formulation, we repeated the analysis including abundances of *Prochlorococcus* on the x-axis. Although doing so increased d values across all biogeochemical conditions, it did not qualitatively affect the relationship as described above (data not shown).

The reduction of *d* reported in Z_l may result from phagotrophy by PPE, by reducing bacterial abundances relative to those of HNF. Feeding experiments in the North Atlantic have demonstrated small plastidic eukaryotes to account for up to 90 % of bacterivory in nutrient-limited waters (Zubkov and Tarran 2008), while laboratory and field studies have demonstrated increased feeding rates specifically in response to P limitation (Christaki et al. 1999; McKie-Krisberg et al. 2015). We also cannot exclude the possibility that the decoupling of bacterial and HNF populations observed at GY reflects increased grazing pressure on HNF by ciliates, which would imply an increase in the importance of top-down processes under nutrient limitation. However, given that ratios of ciliates to bacteria are similar between MA and GY, it does not seem likely that the significant differences in *d* between these sites reflect a change in the interactions between these organisms.

Finally, viruses undoubtedly contribute to the observed variation in bacterial abundances, with previous studies reporting viral lysis to be an equally, if not more important, factor in controlling prokaryote mortality than grazing alone in the surface waters in open ocean, with protistan grazing only becoming dominant in the DCM layer (Lara et al. 2017). Indeed, *d* values were smaller in Z_l (including the DCM) than in Z_u in all three regions investigated in the WTSP. However, the relationship presented by Lara et al (2017) between protist-mediated mortality and

virus-mediated mortality is very large along the 1:1 line, making it difficult to infer how viral lysis might have contributed to the reported differences in *d*.

4.6. Conclusion

Our results demonstrate the distribution of microorganisms in the WTSP to be qualitatively similar to those reported for other highly-stratified oligotrophic regions. The entire transect length was characterized by a two-tier vertical niche partition, with *Prochlorococcus* and *Synechococcus* achieving abundance maxima in the Z_u, and PPE achieving abundance maxima in the Z_l, at depths coincident with the DCM. The strong relationships between N_2 fixation and primary producers demonstrates the central role of N_2 fixation in regulating ecosystem processes in the WTSP, with the influence of biologically fixed nitrogen being exerted across all depths and across all classes of organisms in the study region. At MA and LDB, increases in N_2 fixation rates are accompanied by increased production near the surface, and by increased abundances of *Synechococcus* relative to *Prochlorococcus*. At GY, the marked decrease in N_2 fixation rates is accompanied by greatly reduced phytoplankton abundances, which may translate directly into reduced proportions of HNA bacteria. This shift is coincident with a decoupling of HNF and bacteria populations at GY, suggesting a reduction in the importance of top-down controls in regulating bacteria abundance under nutrient limited conditions. In the lower euphotic zone, these changes may also influence the amount and the quality of nutrients available to PPE communities, influencing both the diversity and vertical distributions of the organisms they comprise.

Table 4.1: Microbial abundance and macronutrient concentrations for each sampling area

Summary of depth-integrated abundances for *Prochlorococcus* (Pro), *Synechococcus* (Syn), picophytoeukaryotes (PPE), and bacteria (Bac), in addition to depth-normalized values of nutrient concentrations (NO_X and PO_4), for different vertical zones (Z_u, Z_l and Z_e) and for individual biogeochemical conditions (cond.: MA, LDB, and GY). $^{*}NO_X = NO_2 + NO_3$.

area	*zone*	*Pro* 10^{11} *cells* m^{-2}	*Syn* 10^{11} *cells* m^{-2}	*PPE* 10^{11} *cells* m^{-2}	*Bac* 10^{11} *cells* m^{-2}	*HNF* 10^{11} *cells* m^{-2}	NO_x* *mmol* m^{-3}	PO_4 *mmol* m^{-3}
MA	Z_u	122 ± 31	3.42 ± 2.44	0.50 ± 0.10	271 ± 73	0.57 ± 0.10	0.02 ± 0.01	0.05 ± 0.04
	Z_l	71 ± 24	0.79 ± 0.42	1.18 ± 0.33	139 ± 26	0.40 ± 0.09	0.43 ± 0.42	0.12 ± 0.05
	Z_e	194 ± 52	4.21 ± 2.53	1.68 ± 0.36	410 ± 97	0.97 ± 0.18	0.22 ± 0.21	0.08 ± 0.03
LDB	Z_u	183 ± 28	16.14 ± 8.64	0.42 ± 0.07	424 ± 108	0.91 ± 0.35	0.01 ± 0.01	0.03 ± 0.01
	Z_l	65 ± 30	2.31 ± 1.13	0.81 ± 0.55	131 ± 40	0.36 ± 0.07	0.02 ± 0.01	0.07 ± 0.02
	Z_e	248 ± 56	18.45 ± 7.68	1.24 ± 0.52	555 ± 141	1.27 ± 0.28	0.02 ± 0.02	0.05 ± 0.01
GY	Z_u	110 ± 9	0.54 ± 0.20	0.63 ± 0.09	290 ± 32	0.41 ± 0.06	0.02 ± 0.01	0.13 ± 0.03
	Z_l	89 ± 10	0.35 ± 0.06	1.26 ± 0.24	164 ± 17	0.43 ± 0.03	0.77 ± 0.42	0.20 ± 0.05
	Z_e	199 ± 9	0.89 ± 0.20	1.89 ± 0.31	455 ± 30	0.85 ± 0.05	0.39 ± 0.20	0.17 ± 0.03

Table 4.2: Two-way ANOVA results for microbial abundance and macronutrient concentrations

Summary table of two-way ANOVA results for parameters analyzed in this study. Row 1 (euphotic layer) tests for significant differences between mean parameter values across different layers of the euphotic layer (Z_u or Z_l) on mean parameter values. Row 2 (condition) tests for significant differences between mean parameter values across different biogeochemical area (MA, LDB, and GY). Row 3 (interaction) tests for differences between mean parameter values across both euphotic layer and biogeochemical condition. Relationships for PRO, SYN, BAC, HNF, NOx, and PO4 calculated from depth-integrated abundances; Relationships for %HNA calculated from raw values. $^{*}NO_X = NO_2 + NO_3$.

		PRO		*SYN*		*PPE*		*BAC*		*%HNA*		*HNF*		*NO_x*		*PO_4*	
	DF	**f**	**p**	**f**	**p**	**f**	**p**	**f**	**p**	**f**	**p**	**f**	**p**	**f**	**p**	**f**	**p**
Layer	1	49.2	< 0.01	23.3	< 0.01	68.1	< 0.01	88.3	< 0.01	5.4	0.02	24.7	< 0.01	42.4	< 0.01	40.6	< 0.01
Area	2	2.9	0.1	25.9	< 0.01	3.2	0.1	4.6	0.2	16.9	< 0.01	6.0	0.01	12.6	< 0.01	113.2	< 0.01
Int.	2	7.0	< 0.01	15.8	< 0.01	0.9	0.4	6.0	0.01	0.7	0.50	11.3	<0.01	11.9	< 0.01	3.01	0.1

Table 4.3: Vertical distributions of HPLC pigments for each biogeochemical area

Average depth-integrated concentrations of 19'-hexanoyloxyfucoxanthin (19'HF), 19'-butanoyloxyfucoxanthin (19'BF) for each biogeochemical area (MA, LDB, and GY). Ratios of these pigments to Chl*a* calculated using HPLC values for TChl*a*.

area	*zone*	*19'BF mg m-3*	*19'HF mg m-3*	*19'BF/Chla*	*19'HF/Chla*	*19'HF/19'BF*
MA	Z_u	0.005 ± 0.007	0.021 ± 0.014	0.037 ± 0.037	0.156 ± 0.060	5.9 ± 2.6
	Z_l	0.040 ± 0.021	0.057 ± 0.031	0.171 ± 0.058	0.240 ± 0.041	1.6 ± 0.7
LDB	Z_u	0.003 ± 0.002	0.022 ± 0.002	0.018 ± 0.016	0.124± 0.049	9.0 ± 3.0
	Z_l	0.011 ± 0.004	0.036 ± 0.006	0.053 ± 0.011	0.186 ± 0.011	3.6 ± 0.9
GY	Z_u	0.003 ± 0.003	0.012 ± 0.007	0.064 ± 0.018	0.254 ± 0.026	4.2 ± 1.2
	Z_l	0.024 ± 0.010	0.037 ± 0.010	0.170 ± 0.051	0.271 ± 0.051	1.7 ± 0.5

Table 4.4: OUTPACE picophytoplankton abundance compared to other data sets

Mean depth-normalized abundances for *Prochlorococcus* (PRO), *Synechococcus* (SYN), picophytoeukaryotes (PPE), and depth- integrated concentration of Chl*a*. %PRO calculated as PRO/(PRO+SYN+PPE). Values for different biogeochemical conditions sampled during OUTPACE transect in bold. Depth normalized values provided to account for differences in the depth of integration between sampling sites (data not shown). OUTPACE Chl*a* values correspond to discrete fluorometric data collected at each station of the transect, with the exception of SD 1-3. Chl*a* values from other datasets correspond to discrete fluorometric or HPLC data.

area	*n*	*Chla* $mg\ m^{-3}$	*Pro* $x\ 10^{11}\ m^3$	*Syn* $x\ 10^{11}\ m^3$	*PPE* $x\ 10^{11}\ m^3$	*% Pro*	*source*
LDB	**3**	**0.34 ± 0.06**	**2.23 ± 0.90**	**0.16 ± 0.07**	**0.01 ± 0.00**	**92**	**This study**
Indian Ocean	28	0.26 ± 0.05	0.12 ± 0.21	0.08 ± 0.06	0.02 ± 0.01	25	JGOFS[1]
MA	**10**	**0.23 ± 0.05**	**1.20 ± 0.34**	**0.02 ± 0.02**	**0.01 ± 0.00**	**97**	**This study**
Arabian Sea	82	0.18 ± 0.07	0.47 ± 0.46	0.23 ± 0.17	0.03 ±0.02	53	JGOFS
N. Pacific	227	0.15 ± 0.02	1.27 ± 0.40	0.01 ± 0.01	0.01 ± 0.00	99	HOT[2]
N. Atlantic	219	0.14 ± 0.06	0.48 ± 0.34	0.07 ± 0.08	0.01 ± 0.01	81	BATS[3] AMT[4]
Eq. Pacific	212	0.14 ± 0.04	1.43 ± 0.54	0.11 ± 0.06	0.05 ± 0.03	93	JGOFS JGOFS AMT
S. Pacific	50	0.12 ± 0.04	1.00 ± 0.34	0.07 ± 0.06	0.03 ± 0.02	93	BiG RAPA[5] BIOSOPE[6] JGOFS JGOFS JGOFS
Med.	16	0.10 ± 0.07	0.33 ± 0.10	0.09 ± 0.09	0.02 ± 0.01	78	JGOFS
GY	**6**	**0.11 ± 0.01**	**1.08 ± 0.10**	**0.02 ± 0.02**	**0.01 ± 0.00**	**99**	**This study**
S. Atlantic	28	0.07 ± 0.04	1.21 ± 0.32	0.02 ± 0.02	0.01 ± 0.01	98	AMT

[1]Joint Global Ocean Flux Study; (Sieger et al. 2005)

[2]Hawaii Ocean Time-series; HOT-Data Organization and Graphical System, http://hahana.soest.hawaii.edu/hot/hot-dogs

[3]**Bermuda Atlantic Time-series, http://bats.bios.edu/bats-data/**
[4]**Atlantic Meridional Transect; British Oceanographic Data Centre, http://www.bodc.ac.uk/**
[5]**C-MORE Data System, http://hahana.soest.hawaii.edu/cmoreDS**
[6]**LEFE-CYBER Database, http://www.obs-vlfr.fr/proof/**

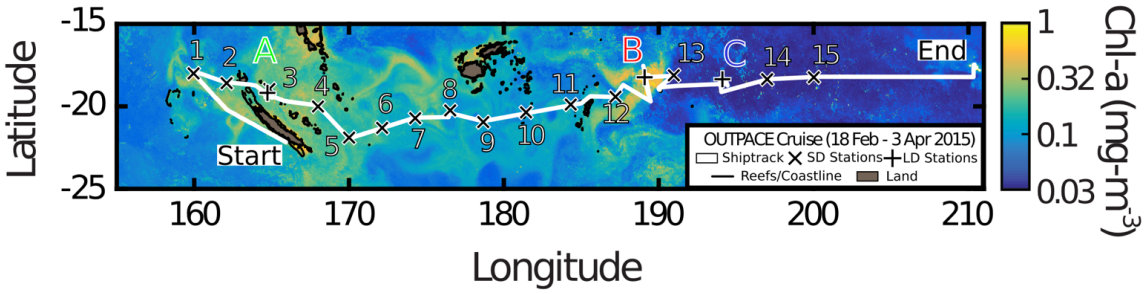

Figure 4.1: Cruise track for OUTPACE

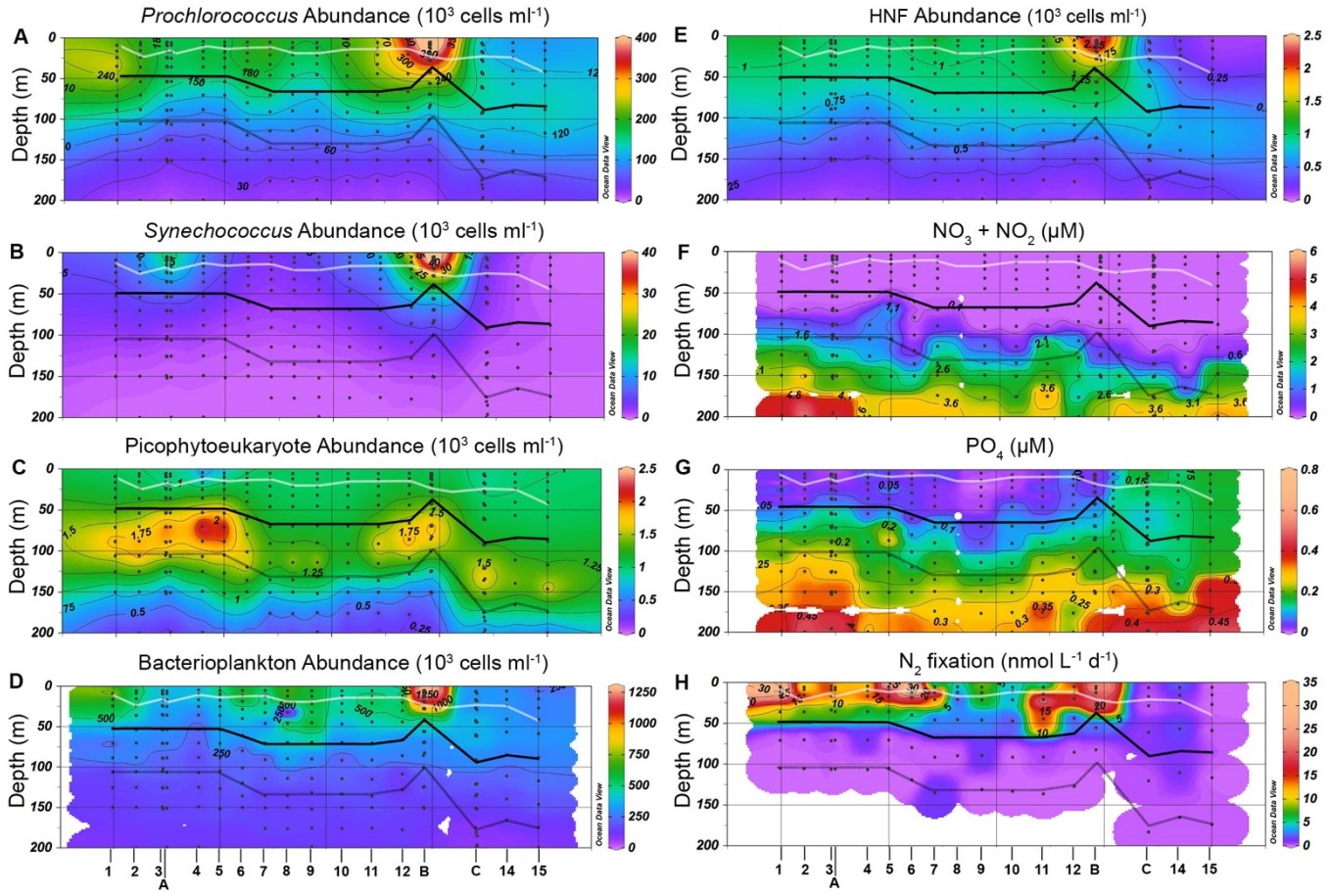

Figure 4.2: Contour plots for measured biological and biogeochemical parameters

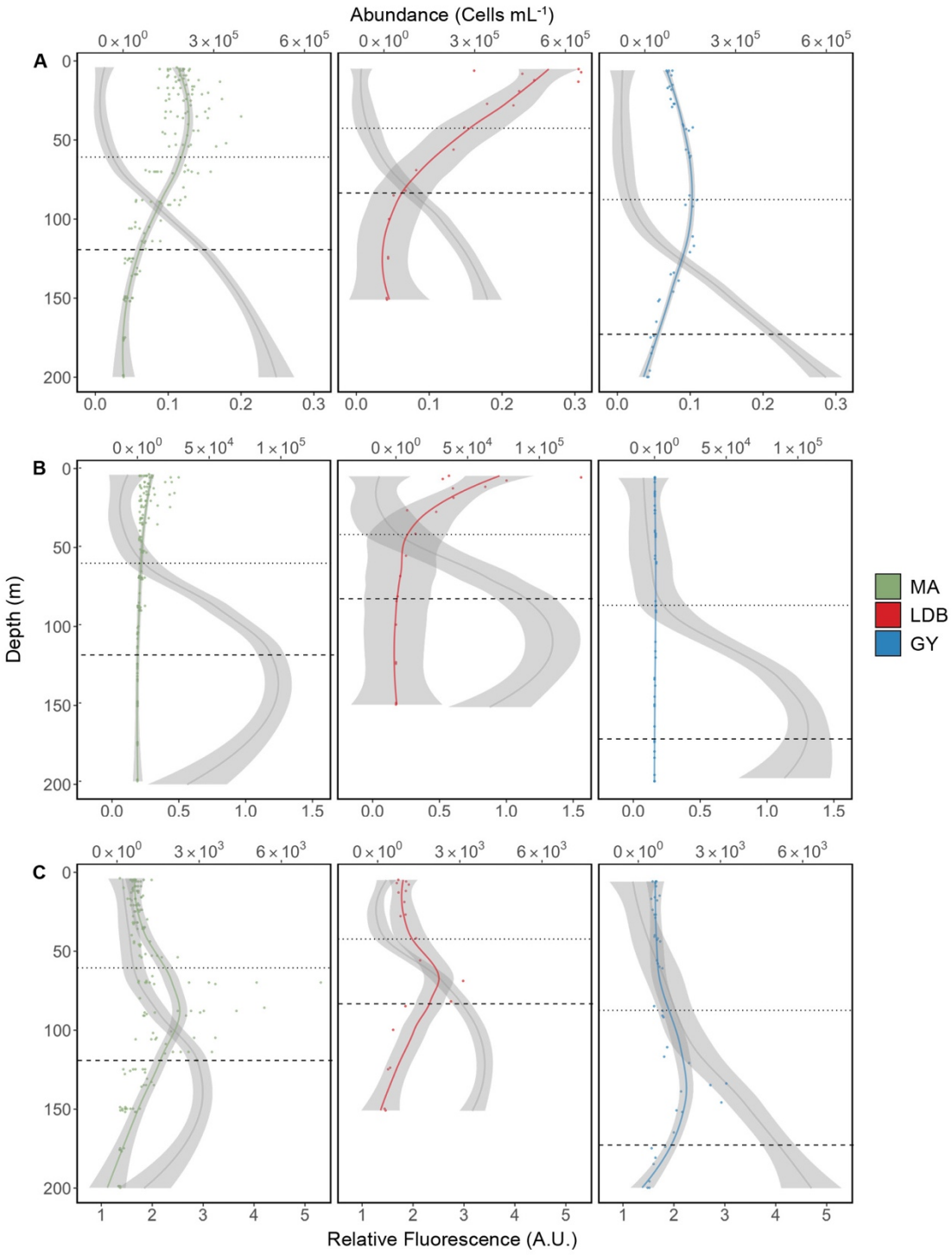

Figure 4.3: Group-specific picophytoplankton abundance and fluorescence profiles

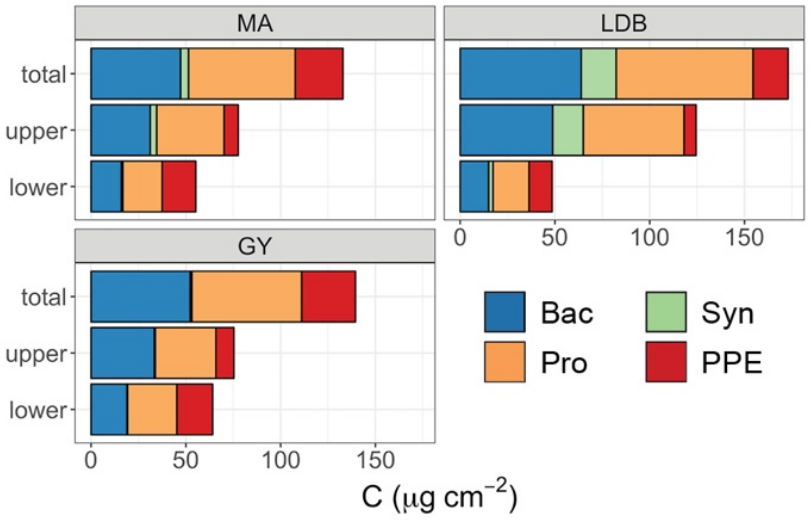

Figure 4.4: Distribution of planktonic biomass for biogeochemical regions sampled

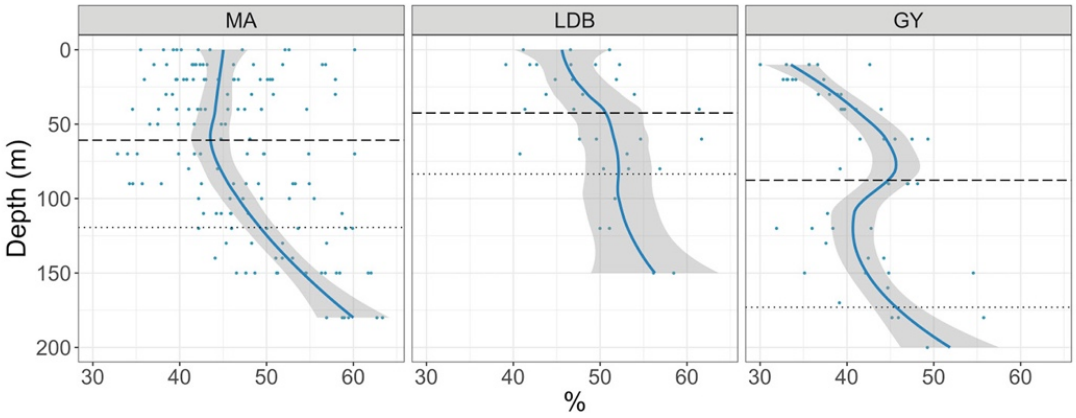

Figure 4.5: Profiles of % HNA for biogeochemical regions sampled

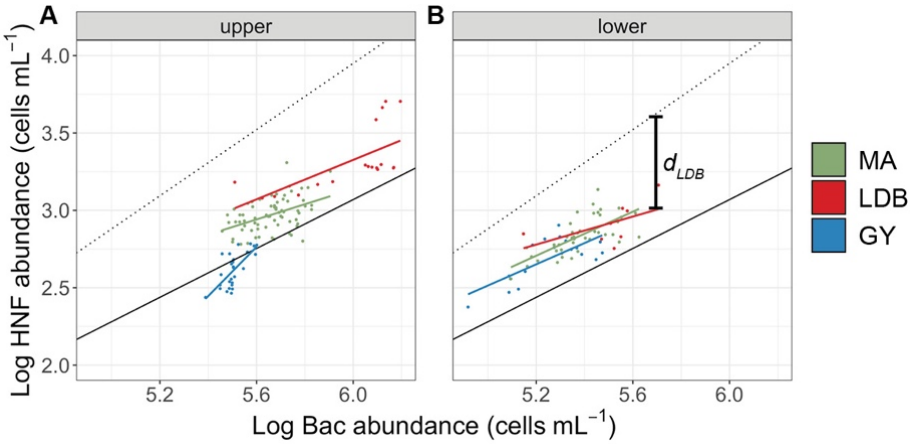

Figure 4.6: Implementation of Gasol model using bacterial and HNF abundancesZ_u

Chapter 5: Experimental Identification and *in silico* Prediction of Bacterivory in Green Algae

5.1. Abstract

While algal phago-mixotrophs play a major role in aquatic microbial food webs, their diversity remains poorly understood. Recent studies have indicated several species of prasinophytes, early diverging green algae, to be able to consume bacteria for nutrition. To further explore the occurrence of phago-mixotrophy in green algae, we conducted feeding experiments with live fluorescently labeled bacteria stained with CellTracker Green CMFDA, heat-killed bacteria stained with 5-(4,6-dichlorotriazin-2-yl) aminofluorescein (DTAF), and magnetic beads. Feeding was detected via microscopy and/or flow cytometry in five strains of prasinophytes when provided with live bacteria: *Pterosperma cristatum* NIES626, *Pyramimonas parkeae* CCMP726, *Pyramimonas parkeae* NIES254, *Nephroselmis pyriformis* RCC618, and *Dolichomastix tenuilepis* CCMP3274. No feeding was detected when heat-killed bacteria or magnetic beads were provided, suggesting a strong preference for live prey in the strains tested. In parallel to experimental assays, green algal bacterivory was investigated using a gene-based prediction model. The predictions agreed with the experimental results and suggested bacterivory potential in additional green algae. Our observations underline the likelihood of widespread occurrence of phago-mixotrophy among green algae, while additionally highlighting potential biases introduced when using prey proxy to evaluate bacterial ingestion by algal cells.

Originally published as Bock, N., Charvet, S., Burns, J., Gyaltshen, Y., Duhamel, S., Kim, E. 2021. Experimental identification and *in silico* prediction of bacterivory in greenalgae. *ISME Journal*. DOI: 10.1038/s41396-021-00899-w

5.2. Introduction

Mixotrophic (or phago-mixotrophic) phytoplankton are microorganisms capable of using both autotrophy (e.g., photosynthesis) and phagotrophy (e.g., bacterivory) to obtain the carbon and nutrients required to support cellular processes. Because of their ability to utilize different trophic modes, mixotrophs are unique and important links in marine microbial food webs. The degree to which they utilize autotrophy versus phagotrophy can have wide-ranging ecological consequences, potentially influencing microbial trophic structure and community stability (Tittel et al. 2003; Jost et al. 2004) in addition to community primary production and carbon cycling (Ward and Follows 2016).

The occurrence of mixotrophy has been long recognized in dinoflagellates (Hansen and Tillmann 2020), haptophytes (Unrein et al. 2014; Anderson et al. 2018), cryptophytes (Lewitus et al. 1991; Yoo et al. 2017) and chrysophyceans (Fenchel 1982; Caron et al. 1990; Holen and Boraas 1995; Rottberger et al. 2013). In addition, there is a growing sense for potential major roles played by green algae in bacterivory in oceanic and other aquatic environments. Bell & Laybourn-Perry (2003), for example, reported ingestion of both fluorescently labeled microspheres and DTAF-labeled bacteria in Arctic strains of *Pyramimonas*. Anderson *et al*. (2018) reported ingestion of DTAF-labeled bacteria by cells in two *Nephroselmis* strains, while McKie-Krisberg *et. al* (McKie-Krisberg and Sanders 2014; McKie-Krisberg et al. 2015) reported ingestion of fluorescent microspheres by both pico- and nano-sized green algae. Paasch (2017) additionally reported stimulated growth in cultures of *Cymbomonas tetramitiformis* inoculated with co-cultured bacteria compared to controls.

Accounting for contributions to total bacterivory made by green algae is crucial to understanding aquatic microbial food web structure and functioning. In particular, prasinophytes—a paraphyletic assemblage of green algae often characterized by the presence of

organic scale covering on the cell body and flagella—are globally distributed and in some cases may account for the majority of phototrophic cells in picophytoplankton size classes (Not et al. 2008; Shi et al. 2009; Rii et al. 2016b) across biogeochemically distinct oceanic regions. Moreover, because the ancestors of Chloroplastida (or Viridiplantae)—comprising green algae and land plants—were presumably phagotrophic, the structure, physiology, and genes that characterize green algal bacterivory may also help illuminate the origin and early history of this highly successful group (Maruyama and Kim 2013). The discovery that *Cymbomonas* has a complex feeding structure that includes the duct system (originally suggested to be involved in transportation of scales to the cell surface) and an acidic spherical compartment where digestion takes place, points to the possibility that other prasinophytes presenting a similar structural organization might also be capable of ingesting bacteria (O'Kelly 1992).

The objective of this study was to explore bacterivory in early diverging green algae by using a combination of experimental assays and gene-based trophic prediction models. Five prasinophyte strains were selected for feeding assays based on morphology, phylogeny, and/or preliminary *in silico* prediction results using an earlier version of the trophic model presented in Burns *et al*. (2018). The selected strains include three pyramimonadophycean strains (*Pterosperma cristatum* NIES 626, *Pyramimonas parkeae* CCMP726, and *Pyramimonas parkeae* NIES 254); a nephroselmidophycean strain (*Nephroselmis pyriformis* RCC 618); and a mamiellophycean strain (*Dolichomastix tenuilepis* CCMP3274). The occurrence of bacterivory in these strains was tested via the detection of ingested fluorescently-labeled bacteria (FLB) via epifluorescence microscopy and flow cytometry. The combined use of these techniques allowed for both the visual confirmation of ingestion via microscopy in addition to the rapid quantification of changes in cellular fluorescence via cytometry (Wilken et al. 2019). Further, we

explored the potential for phagotrophy in diverse green algae using the gene-based trophic model from Burns *et al*. (2018), utilizing a wealth of existing transcriptome and genome data.

5.3. Materials and Methods

5.3.1. Selection of algal strains

All genera of green algae selected for this study are marine flagellates with scales covering the plasma membrane of the cell body and flagella (Inouye et al. 1990; Bhuiyan et al. 2015). Of these, *Pterosperma* and *Pyramimonas* are closely related to *Cymbomonas,* all belonging to the Pyramimonadophyceae; members of which display a duct system or its equivalent that is associated with complex cytoskeletal elements (Inouye et al. 1990; O'Kelly 1992; Adl et al. 2019). In addition, *Pterosperma* was previously shown to endocytose labeled glucans and be positive for acid phosphatase, a conserved component in degradative cellular compartments (Burns et al. 2015). *P. parkeae* CCMP726 and *D. tenuilepis* CCMP3274 were predicted to be mixotrophs using an earlier version of gene-based models as presented in Burns *et al.* (2018). Finally, *N. pyriformis* RCC618 represented a third prasinophyte family, from which others strains (K-0557, K-0556) have been shown to be bacterivorous (Anderson et al. 2018).

5.3.2. Growth of algal strains

Algal cultures were maintained at 17 °C on a 12-hour light/dark cycle. Growth irradiance was approximately 80 μE m^{-2} s^{-1} for all cultures. All algal cultures were monoprotistan and grown in the presence of undefined bacterial flora. For cytometry experiments, aliquots of maintenance cultures in mid-exponential growth were transferred in a 1:10 final dilution to 20 mL of either f/2 (hereafter nutrient replete) or f/20 (hereafter nutrient limited) growth medium. Nutrient replete media was prepared as described in Guillard (1975). Nutrient-limited media was prepared by diluting nutrient replete media 1:10 in ASW. As green algae do not generally require silica, the

silica was not added to the growth media. To determine the growth phase of the cultures, *in vivo* fluorescence measurements were taken at regular intervals using a Turner AquaFluor fluorometer (Turner Designs, San Diego CA). Feeding experiments were conducted when the nutrient replete treatments were in mid to late exponential growth phase.

5.3.3. Preparation of fluorescently labeled bacteria

Fluorescently labeled bacteria (FLB) were prepared from cultures of *Pelagibaca bermudensis* HTCC2601 (Cho and Giovannoni 2006; Thrash et al. 2010), a widespread marine bacterium chosen for its small size (1.2 - 2.3 μm long) relative to the algal strains used. Bacterial cultures were maintained at 27 °C in artificial seawater (ASW) enriched with glucose (0.2 % w/v) and yeast extract (0.5 % w/v). Bacterial cultures were generally in late exponential or stationary growth phase at the time of labeling. To obtain appropriate cell densities for labeling, growth medium was removed from culture flasks using a serological pipette, and the remaining bacterial pellet was resuspended in 1 mL of ASW. *P. bermudensis* cells are non-motile (Cho and Giovannoni 2006) and precipitated naturally during culture growth, eliminating the need for centrifugation to obtain a cell pellet. Triplicate 1 mL subsamples of the cell suspension were labeled with 5 μL SYBR Green I DNA dye (Thermo Fisher Scientific, Waltham MA; Cat. # S785) diluted to a final concentration of 1:10,000, and enumerated via flow cytometry. For the preparation of live labeled FLB the initial cell suspension was diluted in ASW to a final cell abundance of 1×10^9 cells mL^{-1} and a final volume of 1 mL. The remainder of the initial suspension was then used for the preparation of heat-killed FLB.

Live FLB were prepared by labeling bacterial cells with CellTracker Green CMFDA (Thermo Fisher Scientific, Waltham MA; Cat. # C7025): a low toxicity cytoplasmic stain (First et al. 2012). Working solutions of CellTracker Green CMFDA (hereafter CT) were prepared in accordance with manufacturer guidelines and added to one of the dilute cell suspensions to a

final concentration of 10 µM. The CT-labeled FLB suspension (hereafter CT-FLB) was then vortexed for 1 - 2 seconds and incubated for 3 hours in a water bath at 37 °C. Viability tests conducted with *P. bermudensis* indicated bacterial cells were viable after incubation at 37 °C (See Appendix C.1, C.2; Figure C.1). Heat-killed FLB were labeled with 5-(4,6-dichlorotriazin-2-yl) aminofluorescein (5-DTAF, Thermo Fisher Scientific, Waltham MA; Cat. # D16) and prepared as per Sherr *et al*. (Sherr et al. 1987) and Vazquez-Dominguez *et al*. (Vazquez-Dominguez et al. 1999) (hereafter DTAF-FLB). The DTAF-FLB were then vortexed and incubated for two hours in a water bath at 60 °C. *P. bermudensis* viability tests indicated bacterial cells to be no longer viable following incubation at temperatures above 45 °C (See Appendix C.1, C.2; Figure C.1).

To remove excess dye following incubation, each cell suspension was filtered through a 25-mm diameter 0.2-µm porosity polycarbonate filter and rinsed 3× with 1 mL aliquots of ASW. Each filter was transferred to a microcentrifuge tube with 1 mL ASW and vortexed for 30 seconds to detach cells from the filter. The resulting cell suspensions were then transferred to a sterile microcentrifuge tube and stored at 4 °C until the start of experiments. When experiments were conducted more than 12 hours after the preparation of FLB, wash steps were repeated immediately before the start of experiments. All experiments were conducted within 24 hours of the preparation of FLB.

5.3.4. Flow cytometric examination of bacterivory by green algae

All cytometric assays of bacterivory were conducted using a Guava Easycyte flow cytometer (MilliporeSigma, Burlington, MA). For each algal strain tested, triplicate subsamples were prepared from both nutrient replete and nutrient limited culture flasks for two experimental treatments: one inoculated with CT-FLB and one inoculated with DTAF-FLB. To ensure algal cell density did not exceed the maximum number of countable cells by the Guava Easycyte flow

cytometer, initial algal cell densities in experimental cultures were enumerated, with subsamples then being prepared in ASW to a final volume of 500 µL and final algal density of less than 2×10^5 cells mL^{-1}. Treatments were then inoculated with either CT-FLB or DTAF-FLB to final FLB concentrations at 10 % to 50 % of those for unlabeled bacteria. As a negative control (hereafter CT-FLB+PFA), an additional set of triplicate subsamples from both nutrient replete and nutrient limited culture flasks were fixed with 32% electron microscopy grade paraformaldehyde (PFA) to a final concentration of 4% one hour prior to inoculation with CT-FLB. Only one replicate was prepared for *P. parkeae* NIES 254 and CCMP726 treatments inoculated with DTAF-FLB. All treatments were kept in a bench-top incubator, at the same temperature and light intensity as used in the algal growth chamber, for the duration of experiments.

Algal cell populations were identified and enumerated based on chlorophyll red fluorescence and forward scatter signals. Uptake of FLB by algal cells was inferred from changes in the green fluorescence signal of algal cell populations over 3 hours (Figure 5.1; Figure C.2). Green fluorescence was measured cytometrically for each subsample prior to inoculation with FLB to determine the maximum baseline green fluorescence of algal cultures (GF_{prior}; Figure C.2). The duration of experiments was selected based on low ingestion rates in several of the strains used, making it impractical to resolve feeding within intervals less than 1 hour. Subsamples were then analyzed immediately following inoculation with FLB (t_0) and then at 1-hour intervals for 3 hours (t_1, t_2, t_3). An increase in background fluorescence was observed immediately after addition of FLB for all algal cultures and accounted for by measurements at t_0 and then subtracting these values from subsequent times points. For each sample analyzed by flow cytometry, the total algal cell abundance ($cells_{total}$) was recorded. The number of cells with

green fluorescence exceeding GF_{prior}, indicating ingestion of labeled prey, was recorded as $cells_{fed}$. The percentage of $cells_{total}$ having ingested FLB (per_{fed}) was calculated as $(cells_{fed}/cells_{total}) \times 100$. The difference between per_{fed} at t_0 and per_{fed} at t_3 was calculated and expressed as per_{Δ}. Green fluorescence of FLB was measured at t_0 and t_3 in CT, DTAF, and CT+PFA treatments to confirm the stability of FLB fluorescence following fixation.

5.3.5. Epifluorescence microscopic examination of bacterivory by green algae

Algal cultures were incubated in 6×8 multiwell plates (CellTreat Scientific Products, Pepperell, MA) for 15 to 90 min in the algal growth chamber with concentrations of CT-FLB or DTAF-FLB similar to those in the cytometry experiments described above. The prasinophyte *Cymbomonas tetramitiformis* PLY262 and haptophyte *Diacronema lutheri* RCC180, known phago-mixotrophs, were used as general positive controls for FLB ingestion. For strain-specific negative controls to illustrate absence of feeding, the algae were killed in 4% PFA prior to adding CT-FLB. For strain-specific negative controls to check for possible false staining with residual CT stain, the algae were inoculated with the 0.2 μm-filtered supernatant from the last CT-FLB washing step. After incubation, the live algal cells were observed under an epifluorescence microscope (Axiovert 100M, Carl Zeiss, Germany) equipped with filters sets for FITC detection (excitation 470/40, emission 535/40) and chlorophyll detection (excitation 480/30, emission 600LP) to reveal the CellTracker Green or DTAF and natural chlorophyll autofluorescence, respectively. Micrographs were captured with a DP73 camera (Olympus, Japan). The presence of bright green fluorescence inside the algal cell was interpreted as a positive feeding signal. All five prasinophyte strains are characterized by rapid swimming behavior, which, combined with their small cell size, impeded precise counts of ingested bacteria. Furthermore, when exposed to chemical fixatives, including glutaraldehyde, formaldehyde, a mixture of the former two, Lugol's iodine solution and Misky's fixative, the cells

would rapidly (within 2 seconds) lose fluorescence and/or show signs of egesting the fluorescent contents of their vacuoles. Therefore, all observations were carried out on live cells and consisted in qualitative assessment of feeding behavior.

5.3.6. Predictions of phago-mixotrophy using a gene-based model

Predicted protein data from 19 genome and 71 transcriptome assemblies of green algal species, selected as shown below, were tested for phagocytotic potential using a gene-based model formulated by Burns *et al*. (Burns et al. 2018). The majority of the green algal gene sets used in this study originated from the latest version of transcriptome assemblies generated as part of the Marine Microbial Eukaryote Transcriptome Sequencing Project (MMETSP; doi:10.5281/zenodo.3247846) (Leebens-Mack et al. 2019; Wincker 2019) and the Plant 1000 Transcriptome Project (1KP; http://www.onekp.com/) (Keeling et al. 2014; Johnson et al. 2019; Wincker 2019). Coding sequences for all of the transcriptomes were re-annotated using TransDecoder v. 5.5.0 (https://github.com/TransDecoder/TransDecoder). For genome assemblies, whenever gene annotations were not available from the source databases, *de novo* gene prediction was performed with GeneMark-ES v. 4.38 (Besemer 2001) in self-training mode or using models trained on closely related species in case of highly fragmented assemblies. To increase the probability of detecting gene products, protein sets were merged on the level of strains. Completeness of protein sets was assessed with BUSCO v. 4.0.5 (Simão et al. 2015) against the full eukaryote (odb10) lineage. Only the sets with less than 32% missing BUSCO genes were kept for the prediction model, except for the *Cymbomonas* sp. M3265 and *Mesostigma viride* SAG50-1 transcriptomes, which were kept for the analysis despite missing 33.0 and 41.5% of eukaryote orthologs, respectively. We also excluded the *Aphanochaete repens* M2226 protein set showing contaminant sequences matching to a lobose amoeba.

Computational prediction of phagocytotic, photosynthetic, and prototrophic capabilities used the predictTrophicMode tool described in detail in Burns *et al*. (Burns et al. 2018). The code is available at https://github.com/burnsajohn/predictTrophicMode and was run in the default mode after scoring all test genomes and transcriptomes against the hidden Markov models (HMMs) that form the core of the predictTrophicMode tool. Due to broad sampling of eukaryote diversity for the training genomes, the tool is capable of detecting phagocytotic signatures in divergent and novel lineages such as *Mantamonas* (Burns et al. 2018) and the newly discovered *Rhodelphis* (Gawryluk et al. 2019). The predictor is a classifier that was trained by first grouping, by unsupervised clustering, all proteins in 35 complete eukaryote genomes using all vs. all blast followed by MCL clustering. Clusters that satisfied a diversity criterion of containing genes from at least 3 different genomes out of 35 were aligned and converted to HMMs, establishing 14,095 gene models. Individual genes enriched in groups of organisms that share a trophic mode were inferred by grouping the 35 organisms by known trophic capacity for photosynthesis (14 out of 35 organisms), phagocytosis (16 out of 35 organisms), and prototrophy (21 out of 35 organisms). Note that trophic capabilities are not exclusive, organisms could fit into more than one group; see Burns et al. for grouping details (Burns et al. 2018; Table C.1). Enrichment was determined by a significant proportion test (Newcombe 1998). Genes enriched for a trophic capability were clustered by gene ontology (GO) biological processes (BP), and those GO clusters were summarized for each organism using a weighted scoring scheme. The set of the most predictive GO categories was determined using the Boruta algorithm for feature selection (Kursa and Rudnicki 2010). Predictions are called using a probability neural network (PNN), trained using the known trophic capabilities of

the 35 training genomes, which outputs a numerical probability that a new organism will exhibit each capability (Chasset 2013).

Regarding green algae, the tool used three chlorophycean green algal genomes during training, scored as photo-autotrophs, *Chlamydomonas reinhardtii*, *Chlorella variabilis*, and *Volvox carteri*, as well as one prasinophyte, *Cymbomonas tetramitiformis*, scored as a phago-mixotroph. The prediction outputs were visualized by mapping four dimensions with three degrees of freedom (phagocytosis, photosynthesis, prototrophy, and a fourth dependent dimension for absence of each trophic mode) into color space using scripts modified from the R package "pavo" (Maia et al. 2013) and projecting the 4D data onto a 2-dimensional circular Mollweide projection as described in Jimenez *et al*. (2021).

5.3.7. Statistics

One-tailed Student's t-tests were used to identify significant differences between average per_{Δ} values in CT treatments and those in CT+PFA treatments. Shapiro Wilk tests were performed to confirm normality of per_{Δ} values for each treatment. F-tests were performed to evaluate homogeneity of variance between per_{Δ} values for CT and CT+PFA treatments for each strain. One-tailed Welch's t-tests were used in instances where the assumption of homogeneity of variance was not met. A significance threshold of $p \leq 0.05$ was used for all tests. All averages are reported ± standard deviation. All analyses were performed in R Studio (R Core Team 2016).

5.4. Results

5.4.1. Feeding experiments measured by flow cytometry

Results of cytometry feeding experiments are summarized in Figure 5.2. The average per_{Δ} value in the CT+PFA samples (e.g., PFA-killed, negative controls) across all strains tested was 0.4 ± 0.70 % (n = 30) after 3 hours of incubation. The average per_{Δ} value in samples inoculated with heat-killed prey bacteria (DTAF-FLB) across all strains tested was -2.6 ± 3.3 % (n = 22). The

average per_{Δ} value for nutrient replete treatments inoculated with CT-FLB was 4.3 ± 10.2 %. Average per_{Δ} values for nutrient limited treatments inoculated with CT-FLB were 23.6 ± 2.10 % for *N. pyriformis*; 56.79 ± 2.00 % for *P. cristatum*; 64.6 ± 8.23 % for *P. parkeae* CCMP726; 47.7 ± 19.38 % for *P. parkeae* NIES254; and -3.5 ± 0.29 % for *D. tenuilepis*. The negative values for *D. tenuilepis*, which nonetheless was observed to ingest CT-FLB by epifluorescence microscopy, may be attributable to photobleaching of background fluorescence over the course of the incubation (Figure C.4). Per_{Δ} values were significantly greater in nutrient limited treatments inoculated with CT compared to those inoculated with CT+PFA for *N. pyriformis* (Student's t-test, $p = 0.0001$), *P. cristatum* (Welch's t-test, $p = 0.0002$), *P. parkeae* CCMP726 (Welch's t-test, $p = 0.007$), and *P. parkeae* NIES254 (Welch's t-test, $p = 0.02$). Per_{Δ} values for treatments inoculated with CT were not significantly different compared to per_{Δ} values for those inoculated with CT+PFA in nutrient limited *D. tenuilepis* treatments, nor in any nutrient-replete treatments. There were no significant changes in the average green fluorescence of FLB between t_0 and t_3 for CT-FLB, DTAF-FLB or PFA-treated CT-FLB (Student's t-test, $p > 0.05$ for all comparisons). Therefore, feeding results were assumed to not have been biased by changes in the fluorescence of FLB over the course of incubations.

5.4.2. Feeding experiments observed by microscopy

Epifluorescence microscopic observation of algal cultures fed with CT-FLB revealed positive feeding signals in all five tested strains, even *D. tenuilepis* (Figure 5.3; Figures C.5 – C.9), but no evidence of ingestion was observed with the DTAF-FLB (not shown). One or two bright green fluorescent spheres were clearly visible in cells of the two *P. parkeae* strains (Figure 5.3 A–C for NIES254; Figure 5.3 D–F for CCMP726). When *P. cristatum* was inoculated with CT-FLB, we observed a fluorescing sphere that was notably brighter and bigger than the naturally green autofluorescent spot of unknown origin (Figure 5.3 G–I). Some *P. cristatum* cells exhibited two

fluorescent spheres indicative of feeding compartments (not shown). Ingestion of CT-FLB was also observed for *N. pyriformis* (Figure 5.3 J–L) and *D. tenuilepis* by epifluorescence microscopy (Figure 5.3 M–O). Confocal microscopic observations (Appendix C.2) indicated that the green fluorescence was located centrally, lodged between the lobes of the chloroplast and directly under the flagellar pit in *P. parkeae* NIES254 (Figure C10 A) and at or under the base of the flagellar pit for the other green algal strains (Figure C.10 B-D). In contrast to feeding experiments with CT-FLB, we did not find clear indication of magnetic bead ingestion by any of the algal strains (Appendix C.3).

5.4.3. Predictions of phago-mixotrophy using a gene-based model

The predictive model identified 17 out of 90 green algal strains as phago-mixotrophic while most of the green algae clustered as photo-autotrophs and one strain was predicted to be a strict osmotroph, a prototroph that lacks both photosynthesis and phagocytosis (Figure 5.4). The prediction scores represent the probability that a given organism is capable of the function being tested on a scale of 0 to 1. A probability greater than 0.5 is interpreted as evidence that the organism can carry out a given function, with higher scores suggesting increased confidence in the functional prediction. All nine strains of Pyramimonadophyceae surveyed, representing the three genera *Cymbomonas, Pterosperma*, and *Pyramimonas*, were predicted to be phago-mixotrophs with high phagocyte scores (> 0.64). In addition, the two Nephroselmidophyceae strains tested in the model also consistently showed high phagocyte prediction scores, albeit scoring lower than the phago-mixotrophic Pyramimonadophyceae, with 0.519 for *Nephroselmis olivaceae* CCAC0105 and 0.542 for *N. pyriformis* CCMP717. Within the Mamiellophyceae, only two species scored high on phagocyte-generalist predictions: *Dolichomastix tenuilepis* CCMP3274 and M1680 (both with a score greater than 0.9) and *Crustomastix stigmatica* CCMP3273. The other mamiellophyceans with the exception of two *Mantoniella* strains were

predicted as non-phagotrophs with scores below 0.05. *Mantoniella antarctica* and *M. beaufortii* received prediction scores of 0.43 and 0.22, respectively. In addition, two chloropicophycean strains, *Chloropicon primus* CCMP1205 and *Chloropicon laureae* CCMP2175 were predicted as phago-mixotrophs, scoring 0.707 and 0.979, respectively, as well as the basal streptophyte *Mesostigma viride* NIES296 with 0.764. Interestingly, other strains of *Chloropicon* and *Mesostigma* scored comparatively lower in generalist-phagocyte predictions; e.g., *Chloropicon mariensis* RCC998 (0.002) and *C. roscoffensis* CCMP1998, RCC1871 and RCC2335 (0.109, 0.002, 0.007, respectively), and *Mesostigma viride* SAG50-1 (0.006) and NIES995 (0.323). All of the strains mentioned above consistently scored high for photosynthesis and prototrophy predictions (Table C.1). As expected, *Polytomella parva* SAG63-3, a colorless freshwater flagellate with a vestigial plastid (Nakada et al. 2008; Figueroa-Martinez et al. 2015), was predicted as an osmotroph, with a high prototroph score (0.967) but low phagocyte (0.002) and phototroph scores (0.157).

5.5. Discussion

Our experimental results demonstrate the occurrence of bacterivory across all the prasinophyte strains tested, spanning three prasinophyte groups: the Pyramimonadophyceae, the Nephroselmidophyceae and the Mamiellophyceae. The strains, chosen for the feeding experiments in this study, sample a deeper range of diversity within the prasinophytes (Figure 5.5). Among the Pyramimonadophyceae, we chose *P. parkeae* NIES254 and CCMP726, two strains isolated from different sides of the Pacific Ocean, off the coast of Japan and Monterey Bay, respectively. *P. cristatum* NIES626, also isolated in Japan, is the first species of this genus to be confirmed as a bacterivore thus confirming the earlier speculations based on ultrastructural studies (Inouye et al. 1990; O'Kelly 1992; Maruyama and Kim 2013). *Nephroselmis pyriformis*

RCC618 represented the Nephroselmidophyceae (clade III). *N. pyriformis* RCC618 was isolated from the Gulf of Mexico at the Galveston Channel station and is a different strain from the species investigated for bacterivory by Anderson et al. (Anderson et al. 2018), K-0557. *Dolichomastix tenuilepis* CCMP3274, isolated from the Gulf of Naples, Mediterranean Sea, represented the Mamiellophyceae (clade II).

Results from microscopy revealed a green fluorescing compartment within the cells of all five algal strains following inoculation with CT-FLB. Flow cytometry detected significant increases in per_{Δ} for all nutrient limited treatments inoculated with CT-FLB compared to CT+PFA controls, with the exception of *D.tenuilepis*. No significant differences in per_{Δ} values were observed for nutrient replete treatments inoculated with CT-FLB compared to CT+PFA controls. It seems unlikely that significant increases in per_{Δ} were due to factors other than the ingestion of fluorescently labeled prey. The immediate increase in green fluorescence signal observed for all strains with flow cytometry, presumably due to excess extracellular fluorophores in the FLB cell suspensions, was accounted for in per_{Δ} estimates. However, given the limited photostability of fluorescein dyes such as CellTracker CMFDA and DTAF (Johnson 2010), exposure of extracellular fluorophores to light during incubation may have diminished the intensity of background fluorescence over time. Hence, the fluorescence measured likely reflected both increases in algal cell fluorescence due to ingestion of labeled prey and decreases in background fluorescence due to photobleaching of extracellular fluorophores (Figure C.4A). In cases where ingestion rates are low, this phenomenon may result in net negative per_{Δ} values (Figure C.4B), as were observed in some nutrient replete treatments and across all *D. tenuilepis* experiments.

Green algae, including prasinophytes, were traditionally considered as strict phototrophs (Leliaert et al. 2012; Leliaert 2019), despite several earlier studies that suggested the potential phagotrophic behavior in some prasinophytes (Parke and Adams 1961; Thorndsen 1988; O'Kelly 1992; González et al. 1993a; Bell and Laybourn-Parry 2003; Moestrup et al. 2003). More recently, there has been a resurgence of research on bacterivory by green algae, including *Cymbomonas tetramitiformis*, *Micromonas* sp. CCMP2099 (McKie-Krisberg et al. 2015; but see Jimenez et al. 2021), *Pyramimonas tychotreta* and Mantoniella antarctica (McKie-Krisberg et al. 2015), as well as *Nephroselmis pyriformis* (K-0557) and *N. rotunda* (K-0556) (Anderson et al. 2018). The current study further expands the known extent to which bacterivory occurs among green algae. Our gene-based predictive model of phagocytosis, in particular, suggested a potential for bacterivory in several green algal genera, distributed in five distinct green algal sub-lineages (Figure 5.4). The predictions themselves are not determinate; they use gene content from genomic or transcriptomic data to suggest the capacity for phagocytosis. They also do not specify a set of genes that an organism must have to be considered a phagocyte, but rather determine that an organism has a subset of characteristic genes consistent with the capacity for a trait. Therefore, the predictions are most powerful when combined with experimentation. Strains of *Dolichomastix*, *Nephroselmis*, *Pterosperma* and *Pyramimonas* were identified as bacterivores both by *in silico* predictions and empirical evidence. Of the five strains that are newly confirmed to be phagotrophic by laboratory assays in this study, large-scale genetic data currently exist for *Pyramimonas parkeae* CCMP726 and *Dolichomastix tenuilepis* CCMP3274. Therefore, we note that the gene-based model of trophic modes and experimental results concur at least for these two strains.

Furthermore, the model predicted three additional green algal taxa to be phago-mixotrophs, including *Chloropicon primus*, *Crustomastix stigmatica* and *Mesostigma viride*. The dataset used for *Chloropicon primus* CCMP1205 is the first complete genome available for the recently characterized Chloropicophyceae (Turmel et al. 2019). This prediction is particularly surprising in that members of the Chloropicophyceae are small scale-less non-motile cells of 2-3 µm (Lopes dos Santos et al. 2017). Nonetheless, the genome of *Chloropicon primus* contains genetic signatures for the formation of both scales and flagella (Lemieux et al. 2019). Therefore, it is possible that this species has a previously undocumented flagellate life stage, where bacterivory may play a role in nutrition. *Crustomastix stigmatica* is small (3-4 µm long x 2 µm wide) and bears two flagella. Its cell body is not covered with scales, although hairy scales cover the flagella (Zingone et al. 2002). In addition, the prediction of phago-mixotrophy for the genome of the freshwater *M. viride* NIES296 (Liang et al. 2020), an early-diverging streptophyte, supports the hypothesis that bacterivory was present in the common ancestor of Chloroplastida. In terms of characteristic genes in these species, WASH complex proteins, responsible for recycling of membrane proteins from phagosomes back to the cell surface (Buckley et al. 2016), were present in all three datasets (Table C.2). A second phagocytotic hallmark, DNase II (Shpak et al. 2008), is found in both *Chloropicon primus* and *Crustomastix stigmata*, but is absent from *M. viride* (Table C.2).

The model predictions provided contrasting predictions for strains of the same species or genus, such as *Mesostigma* or *Chloropicon*. One strain or species could be predicted to be phago-mixotrophic while another was positioned as photo-autotroph. The contrasting predictions for *Chloropicon* species can in part be seen with the WASH proteins where WASH complex subunits 1 and 4 are detected in all datasets, but subunits 2, 3, and 5 are only detected in the two

strains that received a positive prediction for phagocytosis, *Chloropicon primus* CCMP1205 and *Chloropicon laureae* CCMP2175 (Table C.2). DNase II was detected in all *Chloropicon* strains (Table C.2), regardless of the prediction of phagocytotic capacity, emphasizing the combinatorial nature of the predictions; it is not possible to rely on any one gene or process to make a prediction about phagocytotic capacity. The differential detection of phagocytosis related proteins and resultant contrasting predictions, particularly between strains of the same species, may be attributed to the inherent bias of transcriptomic data in that they reflect protein repertoire expressed in particular physiological conditions. Alternatively, the contrasting interspecific predictions might reflect recent losses of bacterivory in some green algae. For *Mantoniella antarctica*, there is discordance between the model prediction and previous experimental results (Gast et al. 2014). The contrasting claims are reconcilable when considering that the prediction score of 0.43 attained for this alga, abutting the phago-mixotroph threshold, is likely due to incompleteness of the transcriptome data. Indeed, previous predictions of an earlier annotation of this same MMTESP *Mantoniella antarctica* transcriptome provided a phagotroph score of 0.09 (Jimenez et al. 2021), emphasizing the sensitivity of the model to annotation methods. More experimental and sequencing efforts are needed to clarify these issues. Phago-mixotrophy is a spectrum (Mitra et al. 2016) and our model provides avenues to develop new hypotheses on the genetic basis of contrasting mixotrophic behaviors among closely related taxa.

Synthesizing both empirical and *in silico* prediction data, it appears that phago-mixotrophy is broadly preserved among early diverging members of green algae. However, despite their recognized abundance in many marine regions (Moon-van Der Staay et al. 2001; Worden 2006; Not et al. 2008; Kirkham et al. 2013; Rii et al. 2016b), prasinophyte contribution to bacterivory remains conspicuously undetected in most environmental studies (Unrein et al.

2014). The global underestimation of prasinophyte bacterivory might hinge on the class they are represented by in a given environment. Chloropicophyceae tend to be prevalent in the open ocean, such as the oligotrophic South Pacific Subtropical Gyre, where they contribute significantly to primary production (Rii et al. 2016b; Lopes dos Santos et al. 2017). Our *in silico* predictions for *Chloropicon primus* and *C. laureae* suggest that members of this class have the potential to feed on bacteria. Coastal prasinophytes are generally represented by the Mamiellophyceae (Worden 2006; Shi et al. 2009). Our predictions showed phago-mixotrophic potential for some members, including *Dolichomastix tenuilepis* and *Crustomastix stigmatica*, the former of which was also experimentally confirmed to ingest bacteria. However, whether or not other members like *Micromonas* CCMP2099 are capable of consuming bacteria remains debated (McKie-Krisberg and Sanders 2014; Jimenez et al. 2021) and needs to be further investigated. In addition, the limited detection of prasinophyte bacterivory in situ may be attributable to the association of feeding with particular environmental conditions, even for a bacterivorous class, such as the Pyramimonadophyceae for which reports of bacterivory are more frequent from polar systems (Bell and Laybourn-Parry 2003; Gast et al. 2014).

It is also important to consider methodological sources of bias, such as the source of experimental prey. Many protist predators graze selectively (Jürgens and Güde 1994; Jürgens et al. 1999; Suzuki 1999; Van Hannen et al. 1999), with prey size being an important factor. Natural assemblages of marine heterotrophic flagellates and ciliates tend to select larger, actively growing bacterial cells (González et al. 1990, 1993b; Sherr et al. 1992; del Giorgio et al. 1996; Sherr and Sherr 2002), with prey-size in particular affecting the probability of encounter, the capacity for capture and the nutritional value of the prey (Andersen et al. 1986; Fenchel 1986; Epstein and Shiaris 1992). Prey size might also affect the rate at which prey is digested and

decomposed, preventing detection of feeding if ingested prey are digested faster than expected, or if they are egested quickly following ingestion (Pfister and Arndt 1998; Boenigk and Arndt 2002; Pickup et al. 2007; Montagnes et al. 2008). In this study, we observed that *D. tenuilepis* did not ingest a relatively large bacterial prey (*Alteromonas macleodii*) that was successfully consumed by larger strains, including *P. cirstatum* NIES626 and *P. parkeae* NIES254 and CCMP726 (unpublished data).

In addition to the prey size, other factors also play roles in feeding selectivity, such as release of chemical stimuli, prey motility, biochemical composition, and cell surface compounds (Montagnes et al. 2008). Additionally, inert surrogate prey (plastic beads or heat-killed bacteria) have been reported to be less recognizable or ingestible by some protists (González et al. 1993b; Montagnes et al. 2008) such as the phagomixotrophic *Prymnesium* (Legrand et al. 2001) and *Dinobryon* (Caron et al. 1993; Legrand et al. 2001; Montagnes et al. 2008). Therefore, the apparent preference of the green algal strains for the live over the heat-killed FLBs that we observed in this study may be due to significantly reduced ingestion as a result of denaturing of cell structure and biochemistry in the course of the incubations at 60 °C, required for the DTAF labeling protocol. It may also be the case that binding of DTAF fluorophores to the cell surface prevents algal cells from recognizing the DTAF-FLBs as prey. Because Cell Tracker is a cytoplasmic stain, this would not be an issue in CT-FLB. Similar to DTAF-FLB, magnetic beads were not ingested by the green algae although the attachment of a bead to the area near the cell's anterior was occasionally observed for *P. cirstatum* NIES626 (data not shown). In previous experiments with for *P. cirstatum* NIES626, only limited feeding was observed on pHrodo™ bacterial particles, which are chemically fixed, further demonstrating potential preference for live prey in these strains. While we were not able to confirm ingestion of DTAF-FLB or

magnetic beads by any of the green algae, feeding on killed prey might have occurred at a frequency below the detection limits of the methods used in this study (Appendix C.4).

5.6. Conclusions

Mixotrophic small photosynthetic protists enhance carbon transfer within the microbial loop (Fenchel 2008; Ward and Follows 2016). Indeed, as they feed on their competitors for inorganic nutrients these small mixotrophs compete with their own predators for prey biomass (Tittel et al. 2003) and possibly constitute a more nutritious food source for higher trophic levels (Katechakis et al. 2005; Moorthi et al. 2017). In addition, mixotrophic protists also exhibit versatile feeding behaviors (Weisse et al. 2016), which likely affects the nature of the top-down control on prey communities and on carbon cycling. Our results highlight the prevalence of phago-mixotrophy among early-diverging green algae and demonstrate the advantage of using live-stained FLB to assay bacterivory in mixotrophic algae. Contributions of green algal bacterivory to microbial community structure and biogeochemical cycling may have previously been underestimated and require further investigation. Additional research is also needed to experimentally validate *in silico* predictions of phago-mixotrophy in green algae and to characterize the cell structural basis of green algal bacterivory as well as its regulation in response to environmental change.

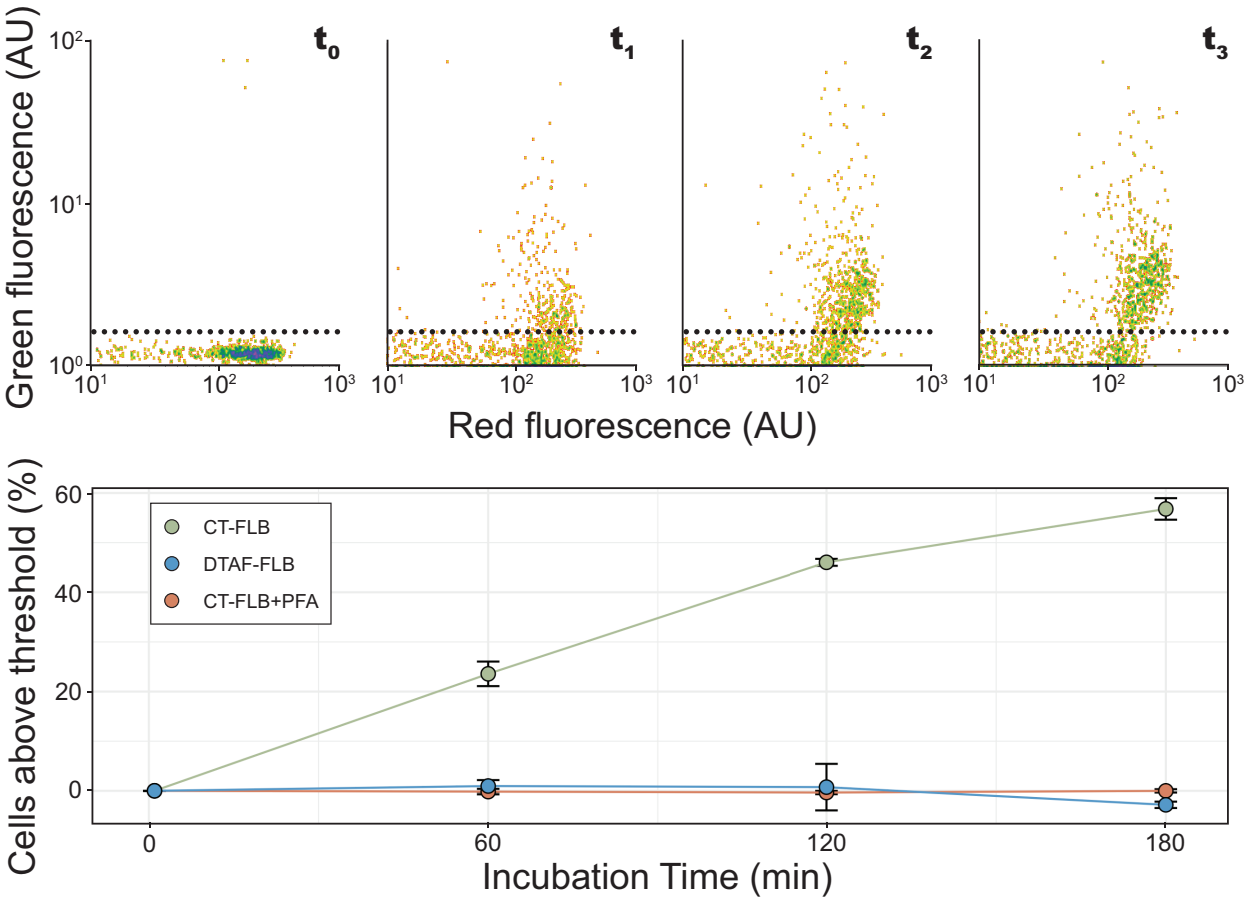

Figure 5.1: **Example of cytometry results for typical feeding experiment**

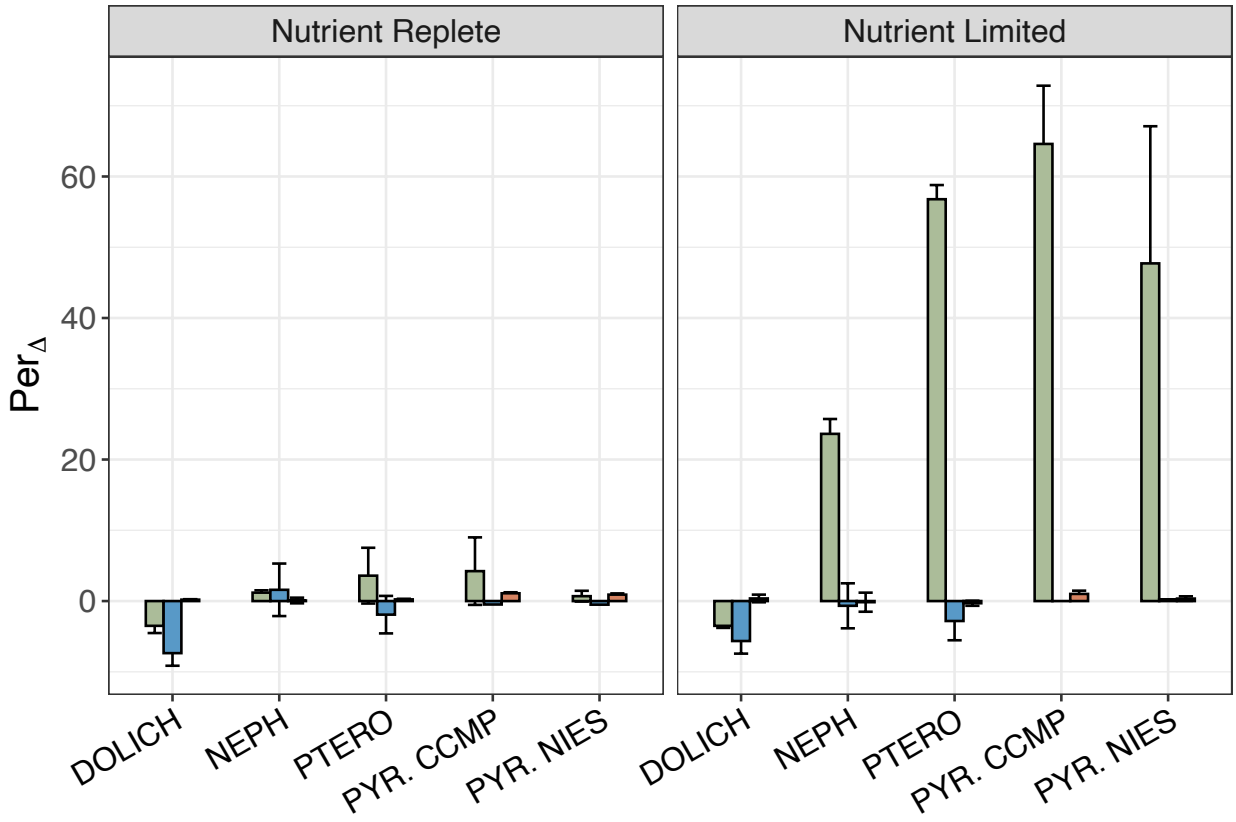

Figure 5.2: **Change in fluorescence of algal cells for individual experimental treatments**

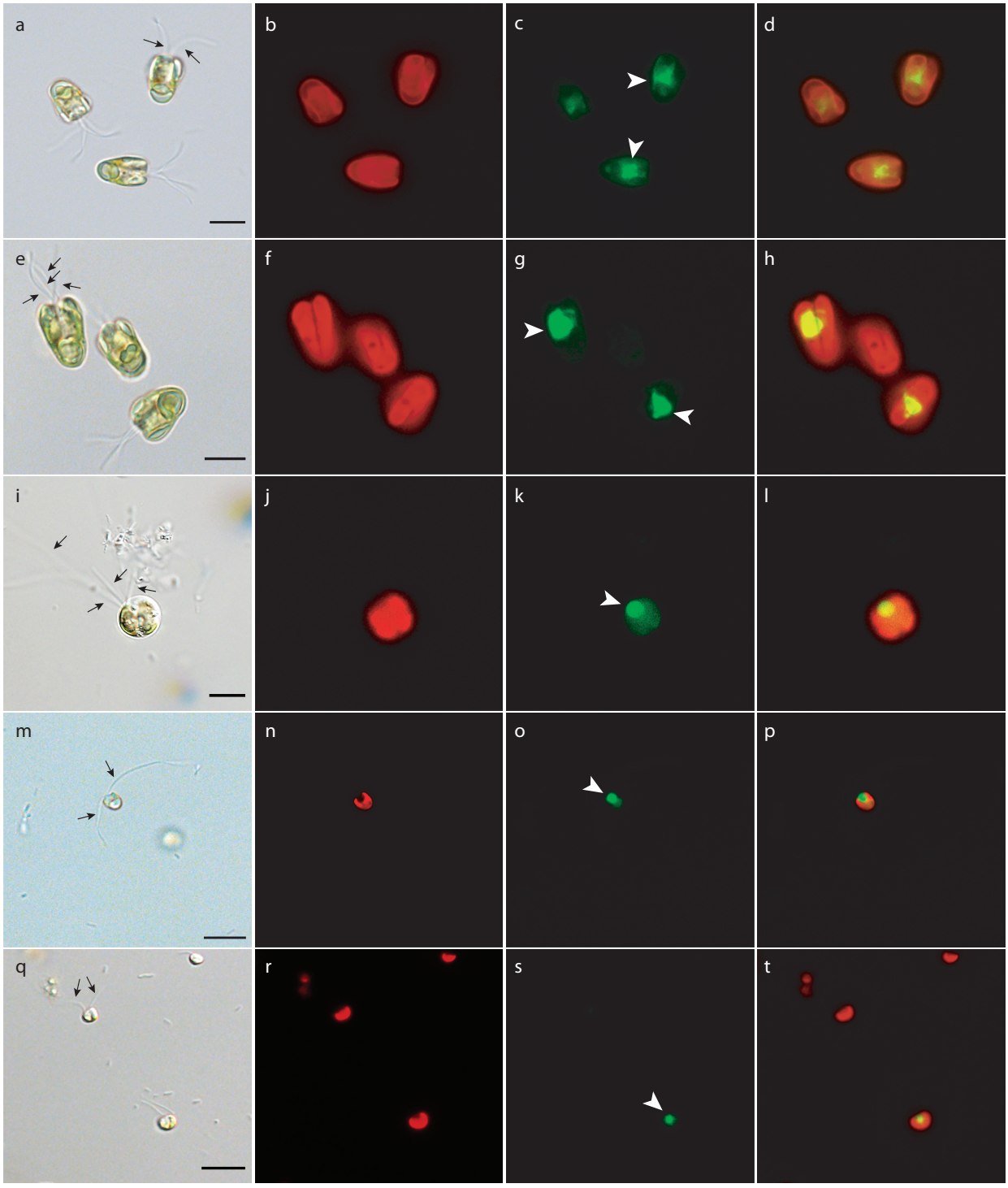

Figure 5.3: Micrographs for algal cultures inoculated with CT-FLB

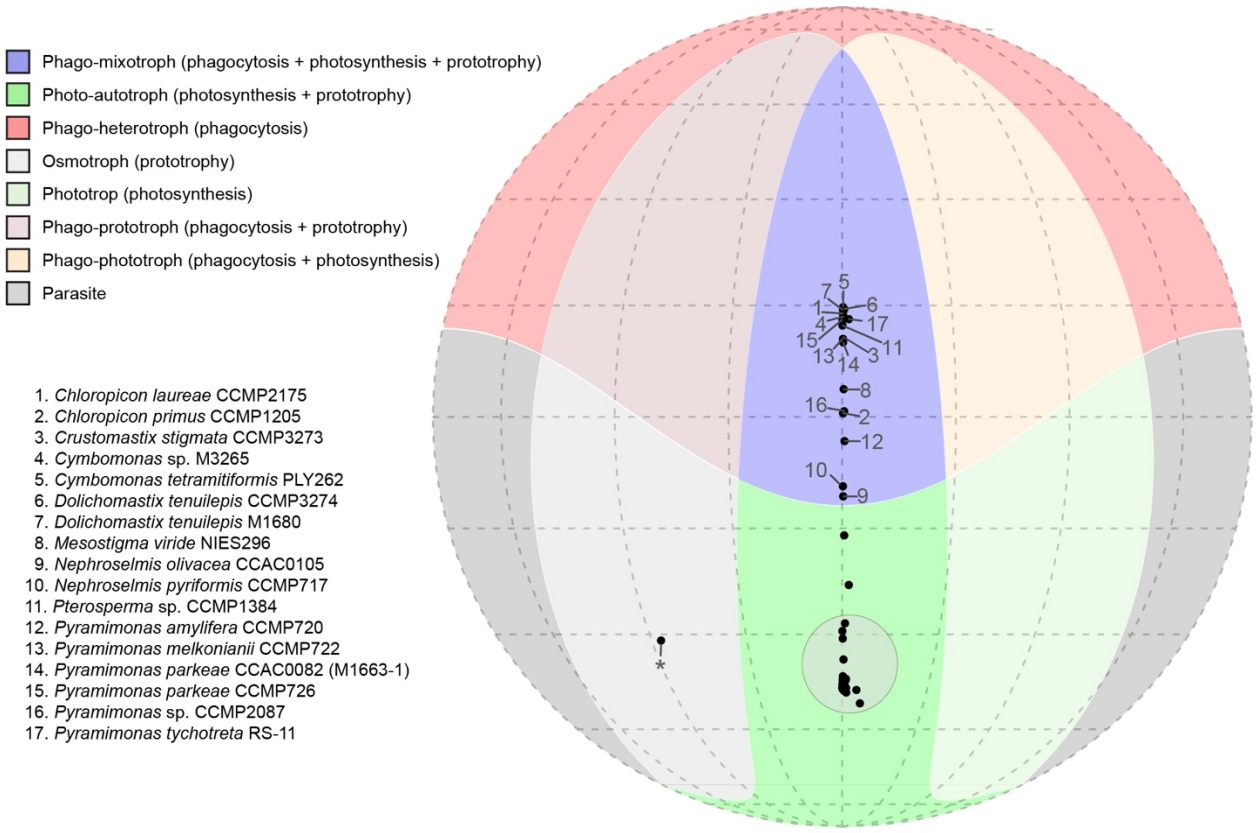

Figure 5.4: Trophic predictions of gene-based model

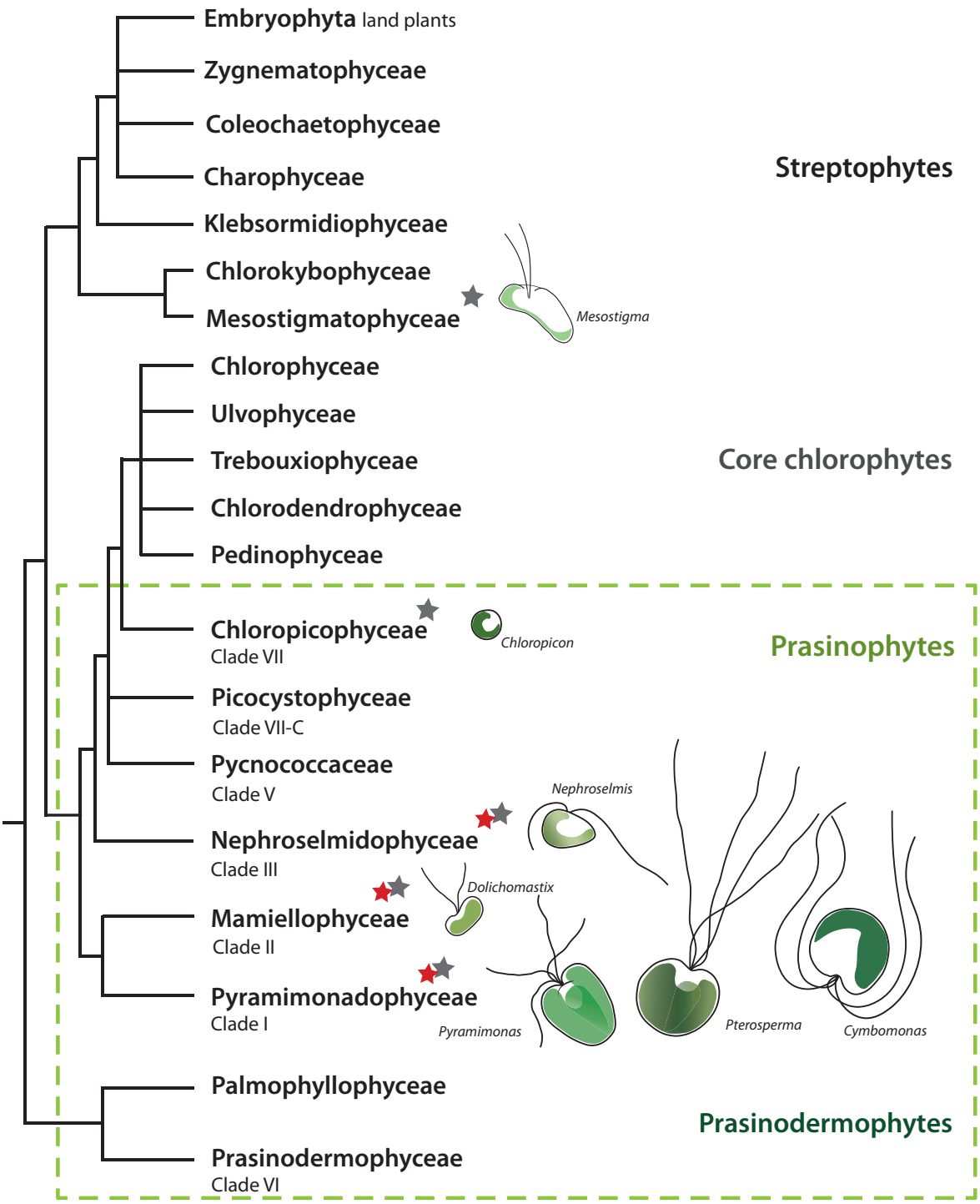

Figure 5.5: **Occurrence of phago-mixotrophy among green algae**

Chapter 6: Conclusion and perspectives

The results presented in this dissertation provide significant contributions to existing knowledge of plankton community structure and function. Chapter 2 introduces a novel biogeographical classification of the global ocean based on subsurface bio-optical measurements from BGC-Argo profiling floats, while also yielding new insight on the environmental factors driving seasonal variability in phytoplankton distributions at the global scale. Chapter 3 introduces a novel methodology for the direct measurement of per-cell Chl*a*, providing the first-ever evaluation of group-specific Chl*a* for *Synechococcus* and PPE based on natural samples and a useful tool for the evaluation of phytoplankton community structure *in situ*. The results of Chapter 4 deliver important new evidence in the ongoing debate over the significance of LNA and HNA subpopulations of heterotrophic bacteria and their influence on the structure of marine food webs and the functioning of the microbial loop, while also highlighting important contributions to phytoplankton biomass by PPE in oligotrophic environments. Finally, Chapter 5 constitutes an advance in the evaluation of bacterivory by algal phagomixotrophs, highlighting important sources of bias that could lead to underestimates of grazing rates and taxonomic diversity of phagomixotrophs in both field and laboratory studies. Together, these results illuminate how environmental variability shapes marine microbial communities at large scales and, moreover, how environmental factors influence microbial ecophysiology at the cellular, population, and community levels.

Aside from providing new knowledge and tools in marine microbial ecology, these results also open new avenues for future research. For example, the biomes identified in Chapter 2 provide a clear route for refining biogeographical provinces based on variance in remote sensing parameters by introducing a geographical constraint that incorporates information on

differences on subsurface variability. Notably, a naive Bayes classifier using annual median $[Chl]_{sat}$, latitude, and longitude as predictors correctly identified biomes in 94% of test cases, suggesting that such an extrapolation may be achievable, especially as more time series data becomes available for constructing the original matrix used for EOF analysis. By incorporating annual median $[Chl]_{sat}$ as a predictor, this biome boundaries would be sensitive to interannual changes in classifications, while still being constrained by latitude and longitude. This would provide basis for a dynamic biogeographical classification, as opposed to static classifications, where biome or province boundaries are fixed in time and space (IOCCG 2009). Within such a framework, applying the method reported in Chapter 3 to obtain direct group-specific measurements of Chl*a* in natural samples could allow for the development of a model for estimating phytoplankton community structure based on bio-optical measurements alone, or for better characterizing variability in the composition and physiology of phytoplankton communities at the depth of the DCM.

The results of Chapter 3 highlight the need for further investigation into how ratios of HNA and LNA bacteria influence food web dynamics and the functioning of the microbial loop, especially given the contradictory results obtained by previous studies (Vaqué et al. 2001; Jochem et al. 2004; Longnecker et al. 2010). While previous studies have demonstrated a size selective feeding in HNF cultures (González et al. 1990), demonstrating a causal chain between nutrient availability, HNA:LNA ratios, and HNF grazing would require experimentally manipulating nutrient concentrations in HNF cultures maintained following continuous culture techniques. Utilizing the method developed in Chapter 5, measuring feeding rates of algal mixotrophs and HNF across a biogeochemical gradient in the field could provide further

evaluation of the hypothesis that decoupling between HNF and bacterial populations is attributable to competition between HNF and phagomixotrophic populations.

Aside from demonstrating the potential for widespread bacterivory in green algae, the prey-selection observed in Chapter 5 highlights the need for further investigation of phagomixotrophy using live labeled prey. In particular, refinement of the methodology utilized in Chapter 5 for field studies will make it possible to 1) evaluate the potential for bias due to prey selectivity *in situ*, 2) the extent to which this bias varies as a function of phytoplankton diversity, and 3) constrain predictions of the role of algal phagomixotrophs in the functioning of marine food webs and in the biogeochemical cycling. And, while our results highlighted a clear increased feeding response to overall nutrient limitation amongst the strains studied, further research is required to evaluate the specific nutrients or nutrient ratios contributing to this response, and the cellular processes modulating phagomixotrophic feeding. This is research I am excited to remain involved in following the completion of my PhD, based on transcriptome data for *Nephroselmis pyriformis* and *Pterosperma cristatum* collected during feeding experiments conducted following those supporting the results presented in Chapter 5.

With the planned deployment of hundreds of new BGC-Argo floats, the ever-increasing spectral resolution provided by sensors deployed aboard remote sensing platforms, and the continued refinement of optical and molecular techniques for studying microbial communities, the next decade stands to bring a transformation in our understanding of global variability in microbial community structure and its role in the functioning of global biogeochemical cycles. Especially given the rapid alteration in global phytoplankton distributions and primary production in response to global climate change (Behrenfeld et al. 2006a; Polovina et al. 2008;

Rousseaux and Gregg 2015), the importance and urgency of this research has never been more important.

References

Adl, S. M., D. Bass, C. E. Lane, and others. 2019. Revisions to the Classification, Nomenclature, and Diversity of Eukaryotes. J. Eukaryot. Microbiol. **66**: 4–119. doi:10.1111/jeu.12691

Álvarez, E., E. Nogueira, and Á. López-Urrutia. 2017. In vivo single-cell fluorescence and size scaling of phytoplankton chlorophyll content. Appl. Environ. Microbiol. **83**: 1–16. doi:10.1128/AEM.03317-16

Andersen, O. K., J. C. Goldman, D. A. Caron, and M. R. Dennett. 1986. Nutrient cycling in a microflagellate food chain: III. Phosphorus dynamics. Mar. Ecol. Prog. Ser. **31**: 47–55.

Anderson, R., S. Charvet, and P. Hansen. 2018. Mixotrophy in chlorophytes and haptophytes - effect of irradiance, macronutrient, micronutrient and vitamin limitation. Front Microbiol **9**: 1704. doi:10.3389/FMICB.2018.01704

Andrade, L., A. M. Gonzalez, C. E. Rezende, M. Suzuki, J. L. Valentin, and R. Paranhos. 2007. Distribution of HNA and LNA bacterial groups in the Southwest Atlantic Ocean. Brazilian J. Microbiol. **38**: 330–336. doi:10.1590/S1517-83822007000200028

Arar, E. J., and G. B. Collins. 1997. U.S. Environmental Protection Agency Method 445.0, In Vitro Determination of Chlorophyll a and Pheophytin a in Marine and Freshwater Algae by Fluorescence, Revision 1.2. U.S. Environmental Protection Agency National Exposure Research Laboratory, Office of Research and Development.

Ardyna, M., L. Lacour, S. Sergi, and others. 2019. Hydrothermal vents trigger massive phytoplankton blooms in the Southern Ocean. Nat. Commun. **10**: 1–8. doi:10.1038/s41467-019-09973-6

Argo. 2000. Argo float data and metadata from Global Data Assembly Centre (Argo GDAC). SEANOE. doi:https://doi.org/10.17882/42182

Azam, F., J. S. Graf, T. Fenchel, J. Field, J. Gray, L. Meyer-Reil, and F. Thingstad. 1983. The Ecological role of water-column microbes in the sea. Mar. Ecol. Prog. Ser. **10**: 257–263. doi:10.3354/meps010257

de Baar, H. J. W., U. Bathmannt, V. Smetacek, B. M. Löscher, and C. Veth. 1995. Importance of iron for plankton blooms and carbon dioxide drawdown in the Southern Ocean. Nature **373**: 412–415. doi:10.1038/373412a0

Barbieux, M., J. Uitz, B. Gentili, and others. 2019. Bio-optical characterization of subsurface chlorophyll maxima in the Mediterranean Sea from a Biogeochemical-Argo float database. Biogeosciences **16**: 1321–1342. doi:10.5194/bg-16-1321-2019

Barboza Tenório, M., C. Dupouy, M. Rodier, and J. Neveux. 2018. *Trichodesmium* and other planktonic cyanobacteria in New Caledonian waters (SW tropical Pacific) during an El Niño episode. Aquat. Microb. Ecol. **81**: 219–241. doi:10.3354/ame01873

Bauer, S., G. L. Hitchcock, and D. B. Olson. 1991. Influence of monsoonally-forced Ekman dynamics upon surface layer depth and plankton biomass distribution in the Arabian Sea. Deep Sea Res. Part A, Oceanogr. Res. Pap. **38**: 531–553. doi:10.1016/0198-0149(91)90062-K

Beckmann, A., and I. Hense. 2007. Beneath the surface: Characteristics of oceanic ecosystems under weak mixing conditions - A theoretical investigation. Prog. Oceanogr. **75**: 771–796. doi:10.1016/j.pocean.2007.09.002

Behrenfeld, M. J., E. Boss, D. A. Siegel, and D. M. Shea. 2005. Carbon-based ocean productivity and phytoplankton physiology from space. Global Biogeochem. Cycles **19**: 1–14. doi:10.1029/2004GB002299

Behrenfeld, M. J., and P. G. Falkowski. 1997. Photosynthetic rates derived from satellite-based chlorophyll concentration. Limnol. Oceanogr. **42**: 1–20. doi:10.4319/lo.1997.42.1.0001

Behrenfeld, M. J., R. T. O'Malley, D. a Siegel, and others. 2006a. Climate-driven trends in contemporary ocean productivity. Nature **444**: 752–755. doi:10.1038/nature05317

Behrenfeld, M. J., T. K. Westberry, E. S. Boss, and others. 2009. Satellite-detected fluorescence reveals global physiology of ocean phytoplankton. Biogeosciences **6**: 779–794. doi:10.5194/bg-6-779-2009

Behrenfeld, M. J., K. Worthington, R. M. Sherrell, F. P. Chavez, P. Strutton, M. McPhaden, and D. M. Shea. 2006b. Controls on tropical Pacific Ocean productivity revealed through nutrient stress diagnostics. Nature **442**: 1025–1028. doi:10.1038/nature05083

Bell, E. M., and J. Laybourn-Parry. 2003. Mixotrophy in the antarctic phytoflagellate, *Pyramimonas gelidicola* (Chlorophyta: Prasinophyceae). J. Phycol. **39**: 644–649. doi:10.1046/j.1529-8817.2003.02152.x

Bellacicco, M., M. Cornec, E. Organelli, and others. 2019. Global Variability of Optical Backscattering by Non-algal particles From a Biogeochemical-Argo Data Set. Geophys. Res. Lett. **46**: 9767–9776. doi:10.1029/2019GL084078

Bellacicco, M., G. Volpe, S. Colella, J. Pitarch, and R. Santoleri. 2016. Influence of photoacclimation on the phytoplankton seasonal cycle in the Mediterranean Sea as seen by satellite. Remote Sens. Environ. **184**: 595–604. doi:10.1016/j.rse.2016.08.004

Berthelot, H., S. Duhamel, S. L'Helguen, J.-F. Maguer, S. Wang, I. Cetinić, and N. Cassar. 2019. NanoSIMS single cell analyses reveal the contrasting nitrogen sources for small phytoplankton. ISME J. **13**: 651–662. doi:10.1038/s41396-018-0285-8

Besemer, J. 2001. GeneMarkS: a self-training method for prediction of gene starts in microbial genomes. Implications for finding sequence motifs in regulatory regions. Nucleic Acids Res. **29**: 2607–2618. doi:10.1093/nar/29.12.2607

Bhuiyan, M. A. H., D. G. Faria, T. Horiguchi, S. D. Sym, and S. Suda. 2015. Taxonomy and

phylogeny of *Pyramimonas vacuolata* sp. nov. (Pyramimonadales, Chlorophyta). Phycologia **54**: 323–332. doi:10.2216/15-004.1

Bittig, H. C., T. Steinhoff, H. Claustre, B. Fiedler, N. L. Williams, R. Sauzède, A. Körtzinger, and J. P. Gattuso. 2018. An alternative to static climatologies: Robust estimation of open ocean CO2 variables and nutrient concentrations from T, S, and O2 data using Bayesian neural networks. Front. Mar. Sci. **5**: 1–30. doi:10.3389/fmars.2018.00328

Blain, S., B. Quéguiner, L. Armand, and others. 2007. Effect of natural iron fertilization on carbon sequestration in the Southern Ocean. Nature **446**: 1070–1074. doi:10.1038/nature05700

Blanchot, J., J. M. André, C. Navarette, and J. Neveux. 1997. Picophytoplankton dynamics in the equatorial Pacific: Diel cycling from flow-cytometer observations. Comptes Rendus l'Academie des Sci. - Ser. III **320**: 925–931. doi:10.1016/S0764-4469(97)80878-5

Blanchot, J., J. M. André, C. Navarette, J. Neveux, and M. H. Radenac. 2001. Picophytoplankton in the equatorial pacific: Vertical distributions in the warm pool and in the high nutrient low chlorophyll conditions. Deep. Res. Part I Oceanogr. Res. Pap. **48**: 297–314. doi:10.1016/S0967-0637(00)00063-7

Blanchot, J., and M. Rodier. 1996. Picophytoplankton abundance and biomass in the western tropical Pacific Ocean during the 1992 El Nino year: Results from flow cytometry. Deep. Res. Part I Oceanogr. Res. Pap. **43**: 877–895. doi:10.1016/0967-0637(96)00026-X

Bock, N., F. Van Wambeke, M. Dion, and S. Duhamel. 2018. Microbial community structure in the western tropical South Pacific. Biogeosciences **15**: 3909–3925. doi:10.5194/bg-15-3909-2018

Boenigk, J., and H. Arndt. 2002. Bacterivory by heterotrophic flagellates: community structure and feeding strategies. Antonie Van Leeuwenhoek **81**: 465–80.

Bonnet, S., H. Berthelot, K. Turk-Kubo, and others. 2016a. Diazotroph derived nitrogen supports diatom growth in the South West Pacific: A quantitative study using nanoSIMS. Limnol. Oceanogr. **61**: 1549–1562. doi:10.1002/lno.10300

Bonnet, S., H. Berthelot, K. Turk-Kubo, S. Fawcett, E. Rahav, S. L'Helguen, and I. Berman-Frank. 2016b. Dynamics of N2 fixation and fate of diazotroph-derived nitrogen in a low-nutrient, low-chlorophyll ecosystem: Results from the VAHINE mesocosm experiment (New Caledonia). Biogeosciences **13**: 2653–2673. doi:10.5194/bg-13-2653-2016

Bonnet, S., M. Caffin, H. Berthelot, and others. 2018. In depth characterization of diazotroph activity across the Western Tropical South Pacific hot spot of N2 fixation. Biogeosciences Discuss. 1–30. doi:10.5194/bg-2017-567

Bouman, H. A., C. Lepère, D. J. Scanlan, and U. Osvaldo. 2012. Phytoplankton community structure in a high-nutrient, low-chlorophyll region of the eastern Pacific Subantarctic region during winter-mixed and summer-stratified conditions. Deep. Res. Part I Oceanogr.

Res. Pap. **69**: 1–11. doi:10.1016/j.dsr.2012.04.008

Boyd, P. W. 2002. Environmental factors controlling phytoplankton processes in the Southern Ocean. J. Phycol. **38**: 844–861. doi:10.1046/j.1529-8817.2002.t01-1-01203.x

Boyd, P. W., and S. C. Doney. 2002. Modelling regional responses by marine pelagic ecosystems to global climate change. Geophys. Res. Lett. **29**: 1806. doi:10.1029/2001gl014130

de Boyer Montégut, C., G. Madec, A. S. Fischer, A. Lazar, and D. Iudicone. 2004. Mixed layer depth over the global ocean: An examination of profile data and a profile-based climatology. J. Geophys. Res. C Ocean. **109**: 1–20. doi:10.1029/2004JC002378

Brewin, R. J. W., X. A. G. Morán, D. E. Raitsos, and others. 2019. Factors regulating the relationship between total and size-Fractionated chlorophyll-a in coastal waters of the Red Sea. Front. Microbiol. **10**: 1–16. doi:10.3389/fmicb.2019.01964

Brewin, R. J. W., S. Sathyendranath, T. Hirata, S. J. Lavender, R. M. Barciela, and N. J. Hardman-Mountford. 2010. A three-component model of phytoplankton size class for the Atlantic Ocean. Ecol. Modell. **221**: 1472–1483. doi:10.1016/j.ecolmodel.2010.02.014

Brewin, R. J. W., S. Sathyendranath, P. K. Lange, and G. Tilstone. 2014. Comparison of two methods to derive the size-structure of natural populations of phytoplankton. Deep Sea Res. Part I Oceanogr. Res. Pap. **85**: 72–79. doi:10.1016/j.dsr.2013.11.007

Briggs, N., M. J. Perry, I. Cetinić, C. Lee, E. D'Asaro, A. M. Gray, and E. Rehm. 2011. High-resolution observations of aggregate flux during a sub-polar North Atlantic spring bloom. Deep. Res. Part I Oceanogr. Res. Pap. **58**: 1031–1039. doi:10.1016/j.dsr.2011.07.007

Brum, J. R., J. C. Ignacio-espinoza, S. Roux, G. Doulcier, S. G. Acinas, A. Alberti, and S. Chaffron. 2015. Ocean Viral Communities. Science **348**: 1261498-1–11. doi:10.1126/science.1261498

Brunet, C., R. Casotti, V. Vantrepotte, F. Corato, and F. Conversano. 2006. Picophytoplankton diversity and photoacclimation in the Strait of Sicily (Mediterranean Sea) in summer. I. Mesoscale variations. Aquat. Microb. Ecol. **44**: 127–141. doi:10.3354/ame044127

Buckley, C. M., N. Gopaldass, C. Bosmani, S. A. Johnston, T. Soldati, R. H. Insall, and J. S. King. 2016. WASH drives early recycling from macropinosomes and phagosomes to maintain surface phagocytic receptors. Proc. Natl. Acad. Sci. **113**: E5906–E5915. doi:10.1073/pnas.1524532113

Burns, J. A., A. Paasch, A. Narechania, and E. Kim. 2015. Comparative genomics of a bacterivorous green alga reveals evolutionary causalities and consequences of phago-mixotrophic mode of nutrition. Genome Biol. Evol. **7**: 3047–3061. doi:10.1093/gbe/evv144

Burns, J. A., A. A. Pittis, and E. Kim. 2018. Gene-based predictive models of trophic modes suggest Asgard archaea are not phagocytotic. Nat. Ecol. Evol. **2**: 697–704.

doi:10.1038/s41559-018-0477-7

Caffin, M., T. Moutin, R. Ann Foster, and others. 2018. N2 fixation as a dominant new N source in the western tropical South Pacific Ocean (OUTPACE cruise). Biogeosciences **15**: 2565–2585. doi:10.5194/bg-15-2565-2018

Calvo-Díaz, A., X. A. G. Morán, and L. Á. Suárez. 2008. Seasonality of picophytoplankton chlorophyll a and biomass in the central Cantabrian Sea, southern Bay of Biscay. J. Mar. Syst. **72**: 271–281. doi:10.1016/j.jmarsys.2007.03.008

Campbell, L., M. R. Landry, J. Constantinou, H. A. Nolla, S. L. Brown, H. Liu, and D. A. Caron. 1998. Response of microbial community structure to environmental forcing in the Arabian Sea. Deep. Res. Part II Top. Stud. Oceanogr. **45**: 2301–2325. doi:Doi 10.1016/S0967-0645(98)00072-1

Campbell, L., and D. Vaulot. 1993. Photosynthetic picoplankton community structure in the subtropical North Pacific Ocean near Hawaii (station ALOHA). Deep. Res. Part I **40**: 2043–2060. doi:10.1016/0967-0637(93)90044-4

Carlson, C., P. del Giorgio, and G. Herndl. 2007. Microbes and the Dissipation of Energy and Respiration: From Cells to Ecosystems. Oceanography **20**: 89–100. doi:10.5670/oceanog.2007.52

Caron, D. A., K. G. Porter, and R. W. Sanders. 1990. Carbon, nitrogen, and phosphorus budgets for the mixotrophic phytoflagellate *Poterioochromonas malhamensis* (Chrysophyceae) during bacterial ingestion. Limnol. Oceanogr. **35**: 433–443. doi:10.4319/lo.1990.35.2.0433

Caron, D. A., R. W. Sanders, E. L. Lim, and others. 1993. Light-Dependent Phagotrophy in the Freshwater Mixotrophic Chrysophyte *Dinobryon cylindricum*. **25**: 93–111.

Casey, J. R., M. W. Lomas, J. Mandecki, and D. E. Walker. 2007. *Prochlorococcus* contributes to new poduction in the Sargasso Sea deep chlorophyll maximum. Geophys. Res. Lett. **34**: 1–5. doi:10.1029/2006GL028725

Cetinić, I., M. J. Perry, N. T. Briggs, E. Kallin, E. A. D'Asaro, and C. M. Lee. 2012. Particulate organic carbon and inherent optical properties during 2008 North Atlantic bloom experiment. J. Geophys. Res. Ocean. **117**. doi:10.1029/2011JC007771

Chai, F., K. S. Johnson, H. Claustre, and others. 2020. Monitoring ocean biogeochemistry with autonomous platforms. Nat. Rev. Earth Environ. **1**: 315–326. doi:10.1038/s43017-020-0053-y

Chasset, P. O. 2013. Probabilistic neural network for the R statistical language. Github. doi:10.5281/zenodo.1483272

Chisholm, S. W., R. J. Olson, E. R. Zettler, R. Goericke, J. B. Waterbury, and N. A. Welschmeyer. 1988. A novel free-living prochlorophyte abundant in the oceanic euphotic zone. Nature **334**: 340–343. doi:10.1038/334340a0

Cho, J.-C., and S. J. Giovannoni. 2006. *Pelagibaca bermudensis gen. nov.*, sp. nov., a novel marine bacterium within the Roseobacter clade in the order Rhodobacterales. Int. J. Syst. Evol. Microbiol. **56**: 855–859. doi:10.1099/ijs.0.64063-0

Chowdhury, M., H. Biswas, A. Mitra, and others. 2021. Southwest monsoon-driven changes in the phytoplankton community structure in the central Arabian Sea (2017–2018): After two decades of JGOFS. Prog. Oceanogr. **197**: 102654. doi:10.1016/j.pocean.2021.102654

Christaki, U. 2001. Nanoflagellate predation on auto- and heterotrophic picoplankton in the oligotrophic Mediterranean Sea. J. Plankton Res. **23**: 1297–1310. doi:10.1093/plankt/23.11.1297

Christaki, U., C. Courties, R. Massana, P. Catala, P. Lebaron, J. M. Gasol, and M. V. Zubkov. 2011. Optimized routine flow cytometric enumeration of heterotrophic flagellates using SYBR Green I. Limnol. Oceanogr. Methods **9**: 329–339. doi:10.4319/lom.2011.9.329

Christaki, U., S. Jacquet, J. R. Dolan, D. Vaulot, and F. Rassoulzadegan. 1999. Growth and grazing on *Prochlorococcus* and *Synechococcus* by two marine ciliates. Limnol. Oceanogr. **44**: 52–61. doi:10.4319/lo.1999.44.1.0052

Claustre, H. 1994. The trophic status of various oceanic provinces as revealed by phytoplankton pigment signatures. Limnol. Oceanogr. **39**: 1206–1210. doi:10.4319/lo.1994.39.5.1206

Claustre, H., K. S. Johnson, and Y. Takeshita. 2020. Observing the global ocean with Biogeochemical-Argo. Ann. Rev. Mar. Sci. **12**: 23–48. doi:10.1146/annurev-marine-010419-010956

Claustre, H., A. Morel, M. Babin, C. Cailliau, D. Marie, J. Marty, D. Tailliez, and D. Vaulot. 1999. Variability in particle attenuation and chlorophyll fluorescence in the tropical Pacific: Scales, patterns, and biogeochemical implications. J. Geophys. Res. **104**: 3401–3422. doi:10.1029/98JC01334

Cornec, M., H. Claustre, A. Mignot, and others. 2021. Deep Chlorophyll Maxima in the Global Ocean: Occurrences, Drivers and Characteristics. Global Biogeochem. Cycles **35**: 1–30. doi:10.1029/2020gb006759

Cuevas, L. A., and C. E. Morales. 2006. Nanoheterotroph grazing on bacteria and cyanobacteria in oxic and suboxic waters in coastal upwelling areas off northern Chile. J. Plankton Res. **28**: 385–397. doi:10.1093/plankt/fbi124

Cullen, J., P. Franks, D. Karl, and A. Longhurst. 2002. Physical influences on marine ecosystem dynamics,.

Cullen, J. J. 2015. Subsurface chlorophyll maximum layers: Enduring enigma or mystery solved? Ann. Rev. Mar. Sci. **7**: 207–239. doi:10.1146/annurev-marine-010213-135111

D'Ortenzio, F., H. Lavigne, F. Besson, and others. 2014. Observing mixed layer depth, nitrate and chlorophyll concentrations in the northwestern Mediterranean: A combined satellite and

NO 3 profiling floats experiment. Geophys. Res. Lett. **41**: 6443–6451. doi:10.1002/2014GL061020

D'Ortenzio, F., and M. Ribera d'Alcalà. 2008. On the trophic regimes of the Mediterranean Sea: a satellite analysis. Biogeosciences Discuss. **5**: 2959–2983. doi:10.5194/bgd-5-2959-2008

Dall'Olmo, G., and K. A. Mork. 2014. Carbon export by small particles in the Norwegian Sea. Geophys. Res. Lett. **41**: 2921–2927. doi:10.1002/2014GL059244

Denis, M., M. Thyssen, V. Martin, B. Manca, and F. Vidussi. 2010. Ultraphytoplankton basin-scale distribution in the eastern Mediterranean Sea in winter: Link to hydrodynamism and nutrients. Biogeosciences **7**: 2227–2244. doi:10.5194/bg-7-2227-2010

Dolan, J. R., A. Gimenez, V. Cornet-barthaux, and A. de Verneil. 2016. Community structure of Tintinnid ciliates of the microzooplankton in the South East Pacific Ocean: Comparison of a high primary productivity with a typical oligotrophic site. J. Eukaryot. Microbiol. **63**: 813–822. doi:10.1111/jeu.12328-4817

Dore, J. E., R. M. Letelier, M. J. Church, R. Lukas, and D. M. Karl. 2008. Summer phytoplankton blooms in the oligotrophic North Pacific Subtropical Gyre: Historical perspective and recent observations. Prog. Oceanogr. **76**: 2–38. doi:10.1016/j.pocean.2007.10.002

Dubinsky, Z., and N. Stambler. 2009. Photoacclimation processes in phytoplankton: Mechanisms, consequences, and applications. Aquat. Microb. Ecol. **56**: 163–176. doi:10.3354/ame01345

Ducklow, H. 2002. Bacterial Production and Biomass in the Oceans, p. 1–47. *In* D. Kirchman [ed.], Microbial Ecology of the Ocean. Wiley.

Ducklow, H. W. 1983. Production and fate of bacteria in the oceans. Bioscience **33**: 494. doi:10.2307/1309138

Ducklow, H. W. 1992. Factors regulating bottom-up control of bacteria biomass in open ocean plankton communities. Arch. Hydrobiol. Beih. Ergebn. Limnol. **37**: 207–217.

Ducklow, H. W., D. A. Purdie, P. J. L. B. Williams, and J. M. Davies. 1986. Bacterioplankton: A sink for carbon in a coastal marine plankton community. Science **232**: 871–873. doi:10.1126/science.232.4752.865

Duhamel, S., K. M. Björkman, J. K. Doggett, and D. M. Karl. 2014. Microbial response to enhanced phosphorus cycling in the North Pacific Subtropical Gyre. Mar. Ecol. Prog. Ser. **504**: 43–58. doi:10.3354/meps10757

Duhamel, S., E. Kim, B. Sprung, and O. R. Anderson. 2019. Small pigmented eukaryotes play a major role in carbon cycling in the P-depleted western subtropical North Atlantic, which may be supported by mixotrophy. Limnol. Oceanogr. **64**: 2424–2440. doi:10.1002/lno.11193

Durand, M. D., R. J. Olson, and S. W. Chisholm. 2001. Phytoplankton population dynamics at the Bermuda Atlantic Time-series station in the Sargasso Sea. Deep. Res. Part II Top. Stud. Oceanogr. **48**: 1983–2003. doi:10.1016/S0967-0645(00)00166-1

Edwards, K. F., M. K. Thomas, C. A. Klausmeier, and E. Litchman. 2015. Light and growth in marine phytoplankton: Allometric, taxonomic, and environmental variation. Limnol. Oceanogr. **60**: 540–552. doi:10.1002/lno.10033

Ellwood, M. J., P. W. Boyd, and P. Sutton. 2008. Winter-time dissolved iron and nutrient distributions in the Subantarctic Zone from 40-52S; 155-160E. Geophys. Res. Lett. **35**: 2–7. doi:10.1029/2008GL033699

Epstein, S., and M. Shiaris. 1992. Size selective grazing of coastal bacterioplankton by natural assemblages of pigmented flagellates, colourless flagellates and ciliates. Microb. Ecol. **23**: 211–225.

Erickson, D. J., J. L. Hernandez, P. Ginoux, W. W. Gregg, C. McClain, and J. Christian. 2003. Atmospheric iron delivery and surface ocean biological activity in the southern ocean and Patagonian region. Geophys. Res. Lett. **30**: 1–4. doi:10.1029/2003GL017241

Evans, G. T., J. S. Parslow, G. T. Evans, and J. S. Parslow. 1985. A model of annual plankton cycles. Deep Sea Res. Part B. Oceanogr. Lit. Rev. **32**: 759. doi:10.1016/0198-0254(85)92902-4

Fenchel, T. 1982. Ecology of heterotrophic microflagellates. II. Bioenergetics and growth. Mar. Ecol. Prog. Ser. **8**: 225–231.

Fenchel, T. 1986. Protistan filter feeding. Prog. Protistol. **1**: 65–113.

Fenchel, T. 2008. The microbial loop - 25 years later. J. Exp. Mar. Bio. Ecol. **366**: 99–103. doi:10.1016/j.jembe.2008.07.013

Fennel, K., and E. Boss. 2003. Subsurface maxima of phytoplankton and chlorophyll: Steady-state solutions from a simple model. Limnol. Oceanogr. **48**: 1521–1534. doi:10.4319/lo.2003.48.4.1521

Ferrier-Pagès, C., and J. P. Gattuso. 1998. Biomass, production and grazing rates of pico- and nanoplankton in coral reef waters (Miyako Island, Japan). Microb. Ecol. **35**: 46–57. doi:10.1007/s002489900059

Field, C. B., M. J. Behrenfeld, J. T. Randerson, C. B. Field, M. J. Behrenfeld, and J. T. Randerson. 1998. Primary production of the biosphere : integrating terrestrial and oceanic components. Science **281**: 237–240.

Figueroa-Martinez, F., A. M. Nedelcu, D. R. Smith, and A. Reyes-Prieto. 2015. When the lights go out: the evolutionary fate of free-living colorless green algae. New Phytol. **206**: 972–982. doi:10.1111/nph.13279

First, M. R., N. Y. Park, M. E. Berrang, R. J. Meinersmann, J. M. Bernhard, R. J. Gast, and J. T. Hollibaugh. 2012. Ciliate ingestion and digestion: Flow cytometric measurements and regrowth of a digestion-resistant Campylobacter jejuni. J. Eukaryot. Microbiol. **59**: 12–19. doi:10.1111/j.1550-7408.2011.00589.x

Fleischer, R. L., P. B. Price, and E. M. Symes. 1964. Novel filter for biological methods. Science 249–250.

Fourrier, M., L. Coppola, H. Claustre, F. D'Ortenzio, R. Sauzède, and J.-P. Gattuso. 2020. A regional neural network approach to estimate water-column nutrient concentrations and carbonate system variables in the Mediterranean Sea: CANYON-MED. Front. Mar. Sci. **7**. doi:10.3389/fmars.2020.00620

Fourrier, M., L. Coppola, H. Claustre, F. D'Ortenzio, R. Sauzède, and J.-P. Gattuso. 2021. Corrigendum: A regional neural network approach to estimate water-column nutrient concentrations and carbonate system variables in the Mediterranean Sea: CANYON-MED. Front. Mar. Sci. **8**. doi:10.3389/fmars.2021.650509

Fuhrman, J. A. 2009. Microbial community structure and its functional implications. Nature **459**: 193–199. doi:10.1038/nature08058

Gasol, J. 1994. A framework for the assessment of top-down vs. bottom-up control of heterotrophic nanoflagellate abundance. Mar. Ecol. Prog. Ser. **113**: 291–300. doi:10.3354/meps113291

Gasol, J. M., C. Pedr, C. Pedrós-Alió, and D. Vaqué. 2002. Regulation of bacterial assemblages in oligotrophic plankton systems: results from experimental and empirical approaches. Antonie Van Leeuwenhoek **81**: 435–452. doi:10.1023/a:1020578418898

Gast, R. J., Z. M. McKie-Krisberg, S. A. Fay, J. M. Rose, and R. W. Sanders. 2014. Antarctic mixotrophic protist abundances by microscopy and molecular methods. FEMS Microbiol. Ecol. **89**: 388–401. doi:10.1111/1574-6941.12334

Gauns, M., M. Madhupratap, N. Ramaiah, R. Jyothibabu, V. Fernandes, J. T. Paul, and S. Prasanna Kumar. 2005. Comparative accounts of biological productivity characteristics and estimates of carbon fluxes in the Arabian Sea and the Bay of Bengal. Deep. Res. Part II Top. Stud. Oceanogr. **52**: 2003–2017. doi:10.1016/j.dsr2.2005.05.009

Gawryluk, R. M. R., D. V. Tikhonenkov, E. Hehenberger, F. Husnik, A. P. Mylnikov, and P. J. Keeling. 2019. Non-photosynthetic predators are sister to red algae. Nature **572**: 240–243. doi:10.1038/s41586-019-1398-6

del Giorgio, P. A., J. M. Gasol, D. Vaqué, P. Mura, S. Agustí, and C. M. Duarte. 1996. Bacterioplankton community structure: Protists control net production and the proportion of active bacteria in a coastal marine community. Limnol. Oceanogr. **41**: 1169–1179. doi:10.4319/lo.1996.41.6.1169

Glover, H. E., B. B. Prezelin, L. Campbell, and others. 1988. A nitrate-dependent *Synechococcus*

bloom in surface Sargasso Sea water. Nature **331**: 161–163. doi:Doi 10.1038/331161a0

Gong, X., W. Jiang, L. Wang, H. Gao, E. Boss, X. Yao, S. J. Kao, and J. Shi. 2017. Analytical solution of the nitracline with the evolution of subsurface chlorophyll maximum in stratified water columns. Biogeosciences **14**: 2371–2386. doi:10.5194/bg-14-2371-2017

González, J. M., B. F. Sherr, and E. B. Sherr. 1993a. Digestive enzyme activity as a quantitative measure of protistan grazing: the acid lysozyme assay for bacterivory. Mar. Ecol. Prog. Ser. **100**: 197–206. doi:10.3354/meps100197

González, J., E. B. Sherr, and B. F. Sherr. 1990. Size-selective grazing on bacteria by natural assemblages of estuarine flagellates and ciliates. Appl. Environ. Microbiol. **56**: 583–589.

González, J., E. B. Sherr, and B. F. Sherr. 1993b. Differential feeding by marine flagellates on growing vs starving bacteria, and on motile vs non-motile bacteria. Mar. Ecol. Prog. Ser. **102**: 257–267.

Graff, J. R., A. J. Milligan, and M. J. Behrenfeld. 2012. The measurement of phytoplankton biomass using flowcytometric sorting and elemental analysis of carbon. Limnol. Oceanogr. Methods **10**: 910–920. doi:10.4319/lom.2012.10.910

Graff, J. R., T. K. Westberry, A. J. Milligan, M. B. Brown, G. Dall'Olmo, K. M. Reifel, and M. J. Behrenfeld. 2016. Photoacclimation of natural phytoplankton communities. Mar. Ecol. Prog. Ser. **542**: 51–62. doi:10.3354/meps11539

Graham, L. E., J. M. Graham, W. L. Wilcox, and M. E. Cook. 2016. Algae, Third. LJLM Press.

Gregg, W. W., and K. L. Carder. 1990. A simple spectral solar irradiance model for cloudless maritime atmospheres. Limnol. Oceanogr. **35**: 1657–1675. doi:10.4319/lo.1990.35.8.1657

Grob, C., O. Ulloa, H. Claustre, Y. Huot, G. Alarcón, and D. Marie. 2007a. Contribution of picoplankton to the total particulate organic carbon concentration in the eastern South Pacific. Biogeosciences **4**: 837–852. doi:10.5194/bg-4-837-2007

Grob, C., O. Ulloa, W. K. W. Li, G. Alarcon, M. Fukasawa, and S. Watanabe. 2007b. Picoplankton abundance and biomass across the eastern South Pacific Ocean along latitude 32.5 degrees S. Mar. Ecol. Prog. Ser. **332**: 53–62. doi:Doi 10.3354/Meps332053

Guidi, L., L. Stemmann, G. A. Jackson, and others. 2009. Effects of phytoplankton community on production, size and export of large Aggregates : A world-ocean analysis. Limnol. Oceanogr. **54**: 1951–1963.

Guillard, R. 1975. Culture of phytoplankton for feeding marine invertebrates, p. 22–60. *In* W.L. Smith and W.H. Chanley [eds.], Culture of Marine Invertebrate Animals. Plenum Press.

Guillou, L., W. Eikrem, M.-J. Chrétiennot-Dinet, F. Le Gall, R. Massana, K. Romari, C. Pedrós-Alió, and D. Vaulot. 2004. Diversity of picoplanktonic Prasinophytes assessed by direct nuclear SSU rDNA sequencing of environmental samples and novel isolates retrieved from

oceanic and coastal marine ecosystems. Protist **155**: 193–214. doi:10.1078/143446104774199592

Gustavs, L., R. Schumann, A. Eggert, and U. Karsten. 2009. In vivo growth fluorometry: Accuracy and limits of microalgal growth rate measurements in ecophysiological investigations. Aquat. Microb. Ecol. **55**: 95–104. doi:10.3354/ame01291

Van Hannen, E. J., M. Veninga, J. Bloem, H. J. Gons, and H. J. Laanbroek. 1999. Genetic changes in the bacterial community structure associated with protistan grazers. Fundam. Appl. Limnol. **145**: 25–38. doi:10.1127/archiv-hydrobiol/145/1999/25

Hansen, P. J., and U. Tillmann. 2020. Mixotrophy among dinoflagellates – prey selection, physiology and ecological imporance, p. 201–260. *In* D.V. Subba Rao [ed.], Dinoflagellates: classification, evolution, physiology and ecological significance. Nova.

Hardman-Mountford, N. J., T. Hirata, K. A. Richardson, and J. Aiken. 2008. An objective methodology for the classification of ecological pattern into biomes and provinces for the pelagic ocean. Remote Sens. Environ. **112**: 3341–3352. doi:10.1016/j.rse.2008.02.016

Hardman-Mountford, N. J., and J. M. McGlade. 2002. 6 Defining ecosystem structure from natural variability: Application of principal components analysis to remotely sensed SST. Large Mar. Ecosyst. **11**: 67–82. doi:10.1016/S1570-0461(02)80028-3

Hastie, T., R. Tibshirani, and J. Friedman. 2009. The Elements of Statistical Learning, Springer New York.

Henson, S. A., R. Sanders, E. Madsen, P. J. Morris, F. Le Moigne, and G. D. Quartly. 2011. A reduced estimate of the strength of the ocean's biological carbon pump. Geophys. Res. Lett. **38**: 10–14. doi:10.1029/2011GL046735

Herbland, A., and B. Voituriez. 1979. Hydrological structure analysis for estimating the primary production in the tropical Atlantic Ocean. J. Mar. Res. **37**: 87–101.

Van Heukelem, L., and C. S. Thomas. 2001. Computer-assisted high performance liquid chromatography method development with applications to the isolation and analysis of phytoplankton pigments. J. Chromatogr. A **23**: 31–49.

Hirata, T., J. Aiken, N. Hardman-mountford, T. J. Smyth, and R. G. Barlow. 2008. An absorption model to determine phytoplankton size classes from satellite ocean colour. Remote Sens. Environ. **112**: 3153–3159. doi:10.1016/j.rse.2008.03.011

Hirata, T., N. J. Hardman-Mountford, R. J. W. Brewin, and others. 2011. Synoptic relationships between surface Chlorophyll-a and diagnostic pigments specific to phytoplankton functional types. Biogeosciences **8**: 311–327. doi:10.5194/bg-8-311-2011

Hobbie, J. E., R. J. Daley, and S. Jasper. 1977. Use of nucleopore filters for counting bacteria by fluorescence microscopy. Appl. Environ. Microbiol. **33**: 1225–1228.

Holen, D. A., and M. E. Boraas. 1995. Mixotrophy in chrysophytes, p. 119–140. *In* Chrysophyte Algae. Cambridge University Press.

Holm-hansen, O., C. Lorenzen, R. Holmes, and J. D. H. Strickland. 1965. Fluorometric determination of chlorophyll. ICES J. Mar. Sci. **30**: 3–15.

Holm-hansen, O., and B. Riemann. 1978. Chlorophyll a determination : Improvements in methodology. Nord. Soc. Oikos **30**: 438–447.

Huang, S., S. Zhang, N. Jiao, and F. Chen. 2015. Marine cyanophages demonstrate biogeographic patterns throughout the global ocean. Appl. Environ. Microbiol. **81**: 441–452. doi:10.1128/AEM.02483-14

Inouye, I., T. Hori, and M. Chihara. 1990. Absolute configuration analysis of the flagellar apparatus of *Pterosperma Cristatum* (Prasinophyceae) and consideration of its phylogenetic position. J. Phycol. **26**: 329–344. doi:10.1111/j.0022-3646.1990.00329.x

IOCCG. 2009. Partition of the ocean into ecological provinces: role of ocean-color radiometry, M. Dowell and T. Platt [eds.].

IUPAC. 1997. Compendium of Chemical Terminology (the "Gold Book"), 2nd ed. A.. McNaught and A. Wilkinson [eds.]. Blackwell Scientific.

James, G., D. Witten, T. Hastie, and R. Tibshirani. 2013. An Introduction to Statistical Learning, Springer New York.

Jardillier, L., M. V Zubkov, J. Pearman, and D. J. Scanlan. 2010. Significant CO2 fixation by small prymnesiophytes in the subtropical and tropical northeast Atlantic Ocean. ISME J. **4**: 1180–1192. doi:10.1038/ismej.2010.36

Jimenez, V., J. A. Burns, F. Le Gall, F. Not, and D. Vaulot. 2021. No evidence of Phago-mixotropy in *Micromonas polaris* (Mamiellophyceae), the Dominant Picophytoplankton Species in the Arctic. J. Phycol. **57**: 435–446. doi:10.1111/jpy.13125

Jochem, F. J., P. J. Lavrentyev, and M. R. First. 2004. Growth and grazing rates of bacteria groups with different apparent DNA content in the Gulf of Mexico. Mar. Biol. **145**: 1213–1225. doi:10.1007/s00227-004-1406-7

Johnson, I. 2010. The Molecular Probes Handbook: A Guide to Fluorescent Probes and Labeling Technologies, 11th ed. Life Technologies Corporation.

Johnson, L. K., H. Alexander, and C. T. Brown. 2019. Re-assembly, quality evaluation, and annotation of 678 microbial eukaryotic reference transcriptomes. Gigascience **8**. doi:10.1093/gigascience/giy158

Johnson, Z. I. 2006. Niche Partitioning Among *Prochlorococcus* Ecotypes Along Ocean-Scale Environmental Gradients. Science **311**: 1737–1740. doi:10.1126/science.1118052

Jost, C., C. A. Lawrence, F. Campolongo, W. Van De Bund, S. Hill, and D. L. DeAngelis. 2004. The effects of mixotrophy on the stability and dynamics of a simple planktonic food web model. Theor. Popul. Biol. **66**: 37–51. doi:10.1016/j.tpb.2004.02.001

Jürgens, K., and H. Güde. 1994. The potential importance of grazing-resistant bacteria in planktonic systems. Mar. Ecol. Prog. Ser. **112**: 169–188. doi:10.3354/meps112169

Jürgens, K., J. Pernthaler, S. Schalla, and R. Amann. 1999. Morphological and compositional changes in a planktonic bacterial community in response to enhanced protozoan grazing. Appl. Environ. Microbiol. **65**: 1241–1250. doi:10.1128/AEM.65.3.1241-1250.1999

Kana, T. M., and P. M. Glibert. 1987. Effect of irradiances up to 2000 μE m^{-2} s^{-1} on marine Synechococcus WH7803-I. Growth, pigmentation, and cell composition. Deep Sea Res. Part A, Oceanogr. Res. Pap. **34**: 479–495. doi:10.1016/0198-0149(87)90001-X

Karl, D. M. 1999. Minireviews: A Sea of Change: Biogeochemical Variability in the North Pacific Subtropical Gyre. Ecosystems **2**: 181–214. doi:10.1007/s100219900068

Karl, D., and L. Proctor. 2007. Foundations of Microbial Oceanography. Oceanography **20**: 16–27. doi:10.5670/oceanog.2007.44

Kashtan, N., S. E. Roggensack, S. Rodrigue, and others. 2014. Single-cell genomics reveals hundreds of coexisting subpopulations in wild *Prochlorococcus*. Science **344**: 416–20. doi:10.1126/science.1248575

Katechakis, A., T. Haseneder, R. Kling, and H. Stibor. 2005. Mixotrophic versus photoautotrophic specialist algae as food for zooplankton: The light : nutrient hypothesis might not hold for mixotrophs. Limnol. Oceanogr. **50**: 1290–1299. doi:10.4319/lo.2005.50.4.1290

Keeling, P. J., F. Burki, H. M. Wilcox, and others. 2014. The Marine Microbial Eukaryote Transcriptome Sequencing Project (MMETSP): Illuminating the functional diversity of eukaryotic life in the oceans through transcriptome sequencing R.G. Roberts [ed.]. PLoS Biol. **12**: e1001889. doi:10.1371/journal.pbio.1001889

Kim, E., B. Sprung, S. Duhamel, C. Filardi, and M. Kyoon Shin. 2016. Oligotrophic lagoons of the South Pacific Ocean are home to a surprising number of novel eukaryotic microorganisms. Environ. Microbiol. **18**: 4549–4563. doi:10.1111/1462-2920.13523

Kirkham, A. R., C. Lepère, L. E. Jardillier, F. Not, H. Bouman, A. Mead, and D. J. Scanlan. 2013. A global perspective on marine photosynthetic picoeukaryote community structure. ISME J. **7**: 922–36. doi:10.1038/ismej.2012.166

Klunder, M. B., P. Laan, R. Middag, H. J. De Baar, and J. C. van Ooijen. 2011. Dissolved iron in the Southern Ocean (Atlantic sector). Deep. Res. Part II Top. Stud. Oceanogr. **58**: 2678–2694. doi:10.1016/j.dsr2.2010.10.042

Kormelink, T. G., G. J. A. Arkesteijn, F. A. Nauwelaers, G. van den Engh, E. N. M. Nolte-'t

Hoen, and M. H. M. Wauben. 2016. Prerequisites for the analysis and sorting of extracellular vesicle subpopulations by high-resolution flow cytometry. Cytom. Part A **89**: 135–147. doi:10.1002/cyto.a.22644

Krug, L. A., T. Platt, S. Sathyendranath, and A. B. Barbosa. 2017. Ocean surface partitioning strategies using ocean colour remote Sensing: A review. Prog. Oceanogr. **155**: 41–53. doi:10.1016/j.pocean.2017.05.013

Kursa, M., and W. Rudnicki. 2010. Feature selection with the Boruta package. J. Stat. Softw. **36**.

Lacour, L., M. Ardyna, K. F. Stec, H. Claustre, L. Prieur, A. Poteau, M. Ribera D'Alcala, and D. Iudicone. 2017. Unexpected winter phytoplankton blooms in the North Atlantic subpolar gyre. Nat. Geosci. **10**: 836–839. doi:10.1038/NGEO3035

Lara, E., I. De Ciències, E. Lara, and others. 2017. Unveiling the role and life strategies of viruses from the surface to the dark ocean. Sci. Adv. doi:10.1126/sciadv.1602565

Lavigne, H., F. D'Ortenzio, C. Migon, and others. 2013. Enhancing the comprehension of mixed layer depth control on the Mediterranean phytoplankton phenology. J. Geophys. Res. Ocean. **118**: 3416–3430. doi:10.1002/jgrc.20251

Lavigne, H., F. Ortenzio, M. Ribera D'Alcala, H. Claustre, R. Sauzede, and M. Gacic. 2015. On the vertical distribution of the chlorophyll a concentration in the Mediterranean Sea : a basin-scale and seasonal approach. Biogeosciences **12**: 5021–5039. doi:10.5194/bg-12-5021-2015

Laws, E. A., P. G. Falkowski, W. O. Smith, H. Ducklow, and J. J. McCarthy. 2000. Temperature effects on export production in the open ocean. Global Biogeochem. Cycles **14**: 1231–1246. doi:10.1029/1999GB001229

Leebens-Mack, J., M. Barker, and E. J. Carpenter. 2019. One thousand plant transcriptomes and the phylogenomics of green plants. Nature **574**: 679–685. doi:10.1038/s41586-019-1693-2

Legrand, C., N. Johansson, G. Johnsen, K. Borsheim, and E. Graneli. 2001. Phagotrophy and toxicity variation in mixotrophic *Prymnesium patelliferum* (Haptophyceae). Limnol. Oceanogr. **46**: 1208–1214.

Leliaert, F. 2019. Green Algae: Chlorophyta and Streptophyta, *In* Reference Module in Life Sciences. Elsevier.

Leliaert, F., D. R. Smith, H. Moreau, M. D. Herron, H. Verbruggen, C. F. Delwiche, and O. De Clerck. 2012. Phylogeny and Molecular Evolution of the Green Algae. CRC. Crit. Rev. Plant Sci. **31**: 1–46. doi:10.1080/07352689.2011.615705

Lemieux, C., C. Otis, and M. Turmel. 2014. Six newly sequenced chloroplast genomes from prasinophyte green algae provide insights into the relationships among prasinophyte lineages and the diversity of streamlined genome architecture in picoplanktonic species. BMC Genomics **15**: 857. doi:10.1186/1471-2164-15-857

Lemieux, C., M. Turmel, C. Otis, and J.-F. Pombert. 2019. A streamlined and predominantly diploid genome in the tiny marine green alga *Chloropicon primus*. Nat. Commun. **10**: 4061. doi:10.1038/s41467-019-12014-x

Lepère, C., D. Vaulot, and D. J. Scanlan. 2009. Photosynthetic picoeukaryote community structure in the South East Pacific Ocean encompassing the most oligotrophic waters on Earth. Environ. Microbiol. **11**: 3105–3117. doi:10.1111/j.1462-2920.2009.02015.x

Letelier, R. M., D. M. Karl, M. R. Abbott, and R. R. Bidigare. 2004. Light driven seasonal patterns of chlorophyll and nitrate in the lower euphotic zone of the North Pacific Subtropical Gyre. Limnol. Oceanogr. **49**: 508–519. doi:10.4319/lo.2004.49.2.0508

Lewitus, A. J., D. A. Caron, and K. R. Miller. 1991. Effect of light and glycerol on the organization of the photosynthetic apparatus in the facultative heterotroph *Pyrenomonas salina* (cryptophyceae). J. Phycol. **27**: 578–587. doi:10.1111/j.0022-3646.1991.00578.x

Li, W. 1995. Composition of ultraphytoplankton in the central North Atlantic. Mar. Ecol. Prog. Ser. **122**: 1–8. doi:10.3354/meps122001

Liang, Z., Y. Geng, C. Ji, and others. 2020. *Mesostigma viride* genome and transcriptome provide insights into the origin and evolution of Streptophyta. Adv. Sci. **7**: 1901850. doi:10.1002/advs.201901850

Loisel, H., and D. Stramski. 2000. Estimation of the inherent optical properties of natural waters from the irradiance attenuation coefficient and reflectance in the presence of Raman scattering. Appl. Opt. **39**: 3001. doi:10.1364/ao.39.003001

Longhurst, A. 1995. Seasonal cycles of pelagic production and consumption. Prog. Oceanogr. **36**: 77–167. doi:10.1016/0079-6611(95)00015-1

Longhurst, A. 2007. Ecological Geography of the Sea, 2nd ed. Elsevier.

Longhurst, A., S. Sathyendranath, T. Platt, and C. Caverhill. 1995. An estimate of global primary production in the ocean from satellite radiometer data. J. Plankton Res. **17**: 1245–1271. doi:10.1093/plankt/17.6.1245

Longnecker, K., M. J. Wilson, E. B. Sherr, and B. F. Sherr. 2010. Effect of top-down control on cell-specific activity and diversity of active marine bacterioplankton. Aquat. Microb. Ecol. **58**: 153–165. doi:10.3354/ame01366

Lönnbro, P., P. Nordenfelt, and H. Tapper. 2008. Isolation of bacteria-containing phagosomes by magnetic selection. BMC Cell Biol. **9**: 35. doi:10.1186/1471-2121-9-35

Lopes dos Santos, A., P. Gourvil, M. Tragin, M. Noël, J. Decelle, S. Romac, and D. Vaulot. 2017. Diversity and oceanic distribution of prasinophytes clade VII , the dominant group of green algae in oceanic waters. ISME J. **11**: 512–528. doi:http://doi.org/10.1038/ismej.2016.210

MacDougall, D., W. B. Crummett, F. J. Amore, and others. 1980. Guidelines for data acquisition and data quality evaluation in environmental chemistry. Anal. Chem. **52**: 2242–2249. doi:10.1021/ac50064a004

Maia, R., C. M. Eliason, P.-P. Bitton, S. M. Doucet, and M. D. Shawkey. 2013. pavo : an R package for the analysis, visualization and organization of spectral data A. Tatem [ed.]. Methods Ecol. Evol. n/a-n/a. doi:10.1111/2041-210X.12069

Marie, D., F. Partensky, D. Vaulot, and C. Brussard. 1999. Enumeration of phytoplankton, bacteria, and viruses in marine samples. Curr. Protoc. Cytom. **10**: 11.11.1-11.11.15. doi:10.1002/0471142956.cy0706s02

Martínez-Pérez, C., W. Mohr, C. R. Löscher, and others. 2016. The small unicellular diazotrophic symbiont, UCYN-A, is a key player in the marine nitrogen cycle. Nat. Microbiol. **1**: 1–7. doi:10.1038/nmicrobiol.2016.163

Maruyama, S., and E. Kim. 2013. A modern descendant of early green algal phagotrophs. Curr. Biol. **23**: 1081–1084. doi:10.1016/j.cub.2013.04.063

Mather, R. L., S. E. Reynolds, G. A. Wolff, and others. 2008. Phosphorus cycling in the North and South Atlantic Ocean subtropical gyres. Nat. Geosci. **1**: 439–443. doi:10.1038/ngeo232

McClain, C. R., S. R. Signorini, and J. R. Christian. 2004. Subtropical gyre variability observed by ocean-color satellites. Deep. Res. Part II Top. Stud. Oceanogr. **51**: 281–301. doi:10.1016/j.dsr2.2003.08.002

Mccreary, J. P., K. E. Kohler, R. R. Hood, and D. B. Olson. 1996. A four-component ecosystem model of biological activity in the Arabian Sea. Prog. Oceanogr. **37**: 193–240. doi:10.1016/S0079-6611(96)00005-5

McKie-Krisberg, Z. M., R. J. Gast, and R. W. Sanders. 2015. Physiological responses of three species of Antarctic mxotrophic phytoflagellates to changes in light and dissolved nutrients. Microb. Ecol. **70**: 21–29. doi:10.1007/s00248-014-0543-x

McKie-Krisberg, Z. M., and R. W. Sanders. 2014. Phagotrophy by the picoeukaryotic green alga Micromonas: implications for Arctic Oceans. ISME J. **8**: 1953–1961. doi:10.1038/ismej.2014.16

Menke, W., and J. Menke. 2016. Environmental Data Analysis with MATLAB, 2nd ed. Elsevier.

Mignot, A., H. Claustre, J. Uitz, A. Poteau, F. D. Ortenzio, and X. Xing. 2014a. Understanding the seasonal dynamics of phytoplankton biomass and the deep chlorophyll maximum in oligotrophic envrionments: A Bio-Argo float investigation. Global Biogeochem. Cycles 856–876. doi:10.1002/2013GB004781.Received

Mignot, A., H. Claustre, J. Uitz, A. Poteau, F. D. Ortenzio, and X. Xing. 2014b. Global Biogeochemical Cycles. AGU. Glob. Biogeochem. cycles 856–876. doi:10.1002/2013GB004781.Received

Mitra, A., K. J. Flynn, U. Tillmann, and others. 2016. Defining planktonic protist functional groups on mechanisms for energy and nutrient acquisition: incorporation of diverse mixotrophic strategies. Protist **167**: 106–120. doi:10.1016/j.protis.2016.01.003

Moestrup, Ø., I. Inouye, and T. Hori. 2003. Ultrastructural studies on *Cymbomonas tetramitiformis* (Prasinophyceae). I. General structure, scale microstructure, and ontogeny. Can. J. Bot. **81**: 657–671. doi:10.1139/b03-055

Montagnes, D., A. Barbosa, J. Boenigk, and others. 2008. Selective feeding behaviour of key free-living protists: avenues for continued study. Aquat. Microb. Ecol. **53**: 83–98.

Moon-van Der Staay, S. Y., R. De Wachter, and D. Vaulot. 2001. Oceanic 18S rDNA sequences from picoplankton reveal unsuspected eukaryotic diversity. **409**: 607–610.

Moore, C. M., M. M. Mills, K. R. Arrigo, and others. 2013. Processes and patterns of oceanic nutrient limitation. Nat. Geosci. **6**: 701–710. doi:10.1038/ngeo1765

Moore, L. R., R. Goericke, and S. W. Chisholm. 1995. Comparative physiology of *Synechococcus* and *Prochlorococcus*: Influence of light and temperature on growth, pigments, fluorescence and absorptive properties. Mar. Ecol. Prog. Ser. **116**: 259–276. doi:10.3354/meps116259

Moore, L. R., A. F. Post, and G. Rocap. 2002. Utilization of different nitrogen sources by the marine cyanobacteria *Prochlorococcus* and *Synechococcus*. Limnol. Oceanogr. **47**: 989–996.

Moore, T. S., J. W. Campbell, and M. D. Dowell. 2009. A class-based approach to characterizing and mapping the uncertainty of the MODIS ocean chlorophyll product. Remote Sens. Environ. **113**: 2424–2430. doi:10.1016/j.rse.2009.07.016

Moore, T. S., J. W. Campbell, and H. Feng. 2001. A fuzzy logic classification scheme for selecting and blending satellite ocean color algorithms. IEEE Trans. Geosci. Remote Sens. **39**: 1764–1776. doi:10.1109/36.942555

Moorthi, S., R. Ptacnik, R. Sanders, R. Fischer, M. Busch, and H. Hillebrand. 2017. The functional role of planktonic mixotrophs in altering seston stoichiometry. Aquat. Microb. Ecol. **79**: 235–245. doi:10.3354/ame01832

Morán, X. A. G., J. M. Gasol, M. C. Pernice, J. F. Mangot, R. Massana, E. Lara, D. Vaqué, and C. M. Duarte. 2017. Temperature regulation of marine heterotrophic prokaryotes increases latitudinally as a breach between bottom-up and top-down controls. Glob. Chang. Biol. **23**: 3956–3964. doi:10.1111/gcb.13730

Morel, A., Y. Ahn, F. Partensky, D. Vaulot, and H. Claustre. 1993. *Prochlorococcus* and *Synechococcus* : A comparative study of their optical properties in relation to their size and pigmentation. J. Mar. Res. **51**: 617–649. doi:10.1357/0022240933223963

Morel, A., and A. Bricaud. 1981. Theoretical results concerning light absorption in a discrete

medium, and application to specific absorption of phytoplankton. Deep Sea Res. **28**: 1375–1393.

Morel, A., B. Gentili, H. Claustre, M. Babin, A. Bricaud, J. Ras, and F. Tièche. 2007. Optical properties of the "clearest" natural waters. Limnol. Oceanogr. **52**: 217–229. doi:10.4319/lo.2007.52.1.0217

Moutin, T., A. Michelangelo Doglioli, A. De Verneil, and S. Bonnet. 2017. Preface: The Oligotrophy to the UlTra-oligotrophy PACific Experiment (OUTPACE cruise, 18 February to 3 April 2015). Biogeosciences **14**: 3207–3220. doi:10.5194/bg-14-3207-2017

Moutin, T., T. Wagener, M. Caffin, and others. 2018. Nutrient availability and the ultimate control of the biological carbon pump in the Western Tropical South Pacific Ocean. Biogeosciences Discuss. 1–41. doi:10.5194/bg-2017-565

Murrell, M. C., and E. M. Lores. 2004. Phytoplankton and zooplankton seasonal dynamics in a subtropical estuary: Importance of cyanobacteria. J. Plankton Res. **26**: 371–382. doi:10.1093/plankt/fbh038

Mustapha, Z. Ben, S. Alvain, C. Jamet, H. Loisel, and D. Dessailly. 2014. Remote Sensing of Environment Automatic classification of water-leaving radiance anomalies from global SeaWiFS imagery : Application to the detection of phytoplankton groups in open ocean waters. Remote Sens. Environ. **146**: 97–112. doi:10.1016/j.rse.2013.08.046

Nakada, T., K. Misawa, and H. Nozaki. 2008. Molecular systematics of *Volvocales* (Chlorophyceae, Chlorophyta) based on exhaustive 18S rRNA phylogenetic analyses. Mol. Phylogenet. Evol. **48**: 281–291. doi:10.1016/j.ympev.2008.03.016

Neveux, J., C. Dupouy, J. Blanchot, A. Le Bouteiller, M. R. Landry, and S. L. Brown. 2003. Diel dynamics of chlorophylls in high-nutrient, low-chlorophyll waters of the equatorial Pacific (180°): Interactions of growth, grazing, physiological responses, and mixing. J. Geophys. Res. C Ocean. **108**: 1–17. doi:10.1029/2000jc000747

Newcombe, R. G. 1998. Interval estimation for the difference between independent proportions: comparison of eleven methods. Stat. Med. **17**: 873–890.

Nishimura, Y., C. Kim, and T. Nagata. 2005. Vertical and seasonal variations of bacterioplankton subgroups with different nucleic acid contents: Possible regulation by phosphorus. Appl. Environ. Microbiol. **71**: 5828–5836. doi:10.1128/AEM.71.10.5828-5836.2005

Not, F., M. Latasa, R. Scharek, and others. 2008. Protistan assemblages across the Indian Ocean, with a specific emphasis on the picoeukaryotes. Deep. Res. Part I Oceanogr. Res. Pap. **55**: 1456–1473. doi:10.1016/j.dsr.2008.06.007

O'Kelly, C. 1992. Flagellar apparatus architecture and the phylogeny of "green" algae: Chlorophytes, Euglenoids, Glaucophytes, p. 315–341. *In* D. Menzel [ed.], The Cytoskeleton of the Algae. CRC Press.

Obata, A., J. Ishizaka, and M. Endoh. 1996. Global verification of critical depth theory for phytoplankton bloom with climatological in situ temperature and satellite ocean color data. J. Geophys. Res. C Ocean. **101**: 20657–20667. doi:10.1029/96JC01734

Olson, R. J., S. W. Chisholm, E. R. Zettler, M. A. Altabet, and J. A. Dusenberry. 1990. Spatial and temporal distributions of prochlorophyte picoplankton in the North Atlantic Ocean. Deep Sea Res. Part A, Oceanogr. Res. Pap. **37**: 1033–1051. doi:10.1016/0198-0149(90)90109-9

Paasch, A. 2017. Physiological and genomic characterization of phagocytosis in green algae. American Museum of Natural History.

Painter, S. C., M. D. Patey, G. A. Tarran, and S. Torres-Valdés. 2014. Picoeukaryote distribution in relation to nitrate uptake in the oceanic nitracline. Aquat. Microb. Ecol. **72**: 195–213. doi:10.3354/ame01695

Parke, M., and I. Adams. 1961. The Pyramimonas-like motile stage of *Halosphaera viridis* Schmitz. Bull. Res. Counc. Isr.

Partensky, F., J. Blanchot, F. Lantoine, J. Neveux, and D. Marie. 1996. Vertical structure of picophytoplankton at different trophic sites of the tropical northeastern Atlantic Ocean. Deep. Res. Part I Oceanogr. Res. Pap. **43**: 1191–1213. doi:10.1016/0967-0637(96)00056-8

Partensky, F., J. Blanchot, and D. Vaulot. 1999. Differential distribution and ecology of Prochlorococcus and Synechococcus in oceanic waters : a review. Bull. l'Institut océanographique **19**: 457–475.

Pérez, V., E. Fernández, E. Marañón, and others. 2006. Vertical distribution of phytoplankton biomass, production and growth in the Atlantic subtropical gyres. Deep Sea Res. Part I Oceanogr. Res. Pap. **53**: 1616–1634. doi:10.1016/j.dsr.2006.07.008

Pernthaler, J. 2005. Predation on prokaryotes in the water column and its ecological implications. Nat. Rev. Microbiol. **3**: 537–546. doi:10.1038/nrmicro1180

Pfister, G., and H. Arndt. 1998. Food selectivity and feeding behaviour in omnivorous filter-feeding ciliates: A case study for Stylonychia. Eur. J. Protistol. **34**: 446–457. doi:10.1016/S0932-4739(98)80013-8

Pickup, Z., R. Pickup, and J. D. Parry. 2007. Growth of Acanthamoeba castellanii and Hartmannella vermiformis on live, heat-killed and DTAF-stained bacterial prey. FEMS Microbiol. Ecol. **61**: 264–272. doi:10.1111/j.1574-6941.2007.00346.x

Pielou, E. 1984. The interpretation of ecological data: a primer on classification and ordination, John Wiley & Sons.

Platt, T., and S. Sathyendranath. 1988. Oceanic primary production: Estimation by remote sensing at local and regional scales. Science **241**: 1613–1620. doi:10.1126/science.241.4873.1613

Platt, T., S. Sathyendranath, C. M. Caverhill, and M. R. Lewis. 1988. Ocean primary production and available light: further algorithms for remote sensing. Deep. Res. Part A. Oceanogr. Res. Pap. **35**: 855–879. doi:10.1016/0198-0149(88)90064-7

Polovina, J. J., E. A. Howell, and M. Abecassis. 2008. Ocean's least productive waters are expanding. Geophys. Res. Lett. **35**: 2–6. doi:10.1029/2007GL031745

Prasanth, R., V. Vijith, V. Thushara, J. V George, and P. N. Vinayachandran. 2021. Deep-Sea Research Part II Processes governing the seasonality of vertical chlorophyll-a distribution in the central Arabian Sea : Bio-Argo observations and ecosystem model simulation.

R Core Team. 2016. R Development Core Team. R A Lang. Environ. Stat. Comput. doi:http://www.R-project.org

Ras, J., H. Claustre, and J. Uitz. 2008. Spatial variability of phytoplankton pigment distributions in the Subtropical South Pacific Ocean: comparison between in situ and predicted data. Biogeosciences **5**: 353–369.

Ravichandran, M., M. S. Girishkumar, and S. Riser. 2012. Observed variability of chlorophyll-a using Argo profiling floats in the southeastern Arabian Sea. Deep. Res. Part I Oceanogr. Res. Pap. **65**: 15–25. doi:10.1016/j.dsr.2012.03.003

Rembauville, M., N. Briggs, M. Ardyna, and others. 2017. Plankton Assemblage Estimated with BGC-Argo Floats in the Southern Ocean: Implications for Seasonal Successions and Particle Export. J. Geophys. Res. Ocean. **122**: 8278–8292. doi:10.1002/2017JC013067

Rii, Y., D. Karl, and M. Church. 2016a. Temporal and vertical variability in picophytoplankton primary productivity in the North Pacific Subtropical Gyre. Mar. Ecol. Prog. Ser. **562**: 1–18. doi:10.3354/meps11954

Rii, Y. M., S. Duhamel, R. R. Bidigare, D. M. Karl, D. J. Repeta, and M. J. Church. 2016b. Diversity and productivity of photosynthetic picoeukaryotes in biogeochemically distinct regions of the South East Pacific Ocean. Limnol. Oceanogr. **61**: 806–824. doi:10.1002/lno.10255

Roemmich, D., M. H. Alford, H. Claustre, and others. 2019. On the future of Argo: A global, full-depth, multi-disciplinary array. Front. Mar. Sci. **6**: 439. doi:10.3389/fmars.2019.00439

Rottberger, J., A. Gruber, J. Boenigk, and P. Kroth. 2013. Influence of nutrients and light on autotrophic, mixotrophic and heterotrophic freshwater chrysophytes. Aquat. Microb. Ecol. **71**: 179–191. doi:10.3354/ame01662

Roullier, F., L. Berline, L. Guidi, X. Durrieu De Madron, M. Picheral, A. Sciandra, S. Pesant, and L. Stemmann. 2014. Particle size distribution and estimated carbon flux across the Arabian Sea oxygen minimum zone. Biogeosciences **11**: 4541–4557. doi:10.5194/bg-11-4541-2014

Rousseaux, C., and W. Gregg. 2015. Recent decadal trends in global phytoplankton composition.

Global Biogeochem. Cycles **20**: 1674–1688. doi:10.1002/2015GB005139

Rousseeuw, P. J. 1987. Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. J. Comput. Appl. Math. **20**: 53–65. doi:10.1016/0377-0427(87)90125-7

Sarmiento, J. L., R. Slater, R. Barber, and others. 2004. Response of ocean ecosystems to climate warming. Global Biogeochem. Cycles **18**: 6–8. doi:10.1029/2003GB002134

Sato, M., S. Takeda, and K. Furuya. 2006. Effects of long-term sample preservation on flow cytometric analysis of natural populations of pico-and nanophytoplankton. J. Oceanogr. **62**: 903–908. doi:10.1007/s10872-006-0107-x

Sauzède, R., H. C. Bittig, H. Claustre, O. P. de Fommervault, J. P. Gattuso, L. Legendre, and K. S. Johnson. 2017. Estimates of water-column nutrient concentrations and carbonate system parameters in the global ocean: A novel approach based on neural networks. Front. Mar. Sci. **4**: 1–18. doi:10.3389/fmars.2017.00128

Schine, C. M. S., A. C. Alderkamp, G. van Dijken, and others. 2021. Massive Southern Ocean phytoplankton bloom fed by iron of possible hydrothermal origin. Nat. Commun. **12**: 1–11. doi:10.1038/s41467-021-21339-5

Schlesinger, W. H., and E. S. Bernhardt. 2013. Biogeochemistry, 4th ed. Elsevier.

Schlitzer, R. 2017. Ocean Data View.

Schmechtig, C., A. Poteau, H. Claustre, F. D'Ortenzio, and E. Boss. 2015. Processing Bio-Argo chlorophyll-a concentration at the DAC Level. Argo Data Manag. 1–22.

Schmechtig, C., A. Poteau, H. Claustre, F. D'Ortenzio, G. Dall'Olmo, and E. S. Boss. 2018a. Processing BGC-Argo particle backscattering at the DAC level. 1–15.

Schmechtig, C., A. Poteau, H. Claustre, F. D'Ortenzio, G. Dall'Olmo, and E. S. Boss. 2018b. Bio-Argo quality control manual for Chlorophyll-A concentration. 1–16. doi:http://dx.doi.org/10.13155/39459

Sedwick, P. N., A. R. Bowie, and T. W. Trull. 2008. Dissolved iron in the Australian sector of the Southern Ocean (CLIVAR SR3 section): Meridional and seasonal trends. Deep. Res. Part I Oceanogr. Res. Pap. **55**: 911–925. doi:10.1016/j.dsr.2008.03.011

Sharples, J., C. M. Moore, T. P. Rippeth, P. M. Holligan, D. J. Hydes, N. R. Fisher, and J. H. Simpson. 2001. Phytoplankton distribution and survival in the thermocline. Limnol. Oceanogr. **46**: 486–496. doi:10.4319/lo.2001.46.3.0486

Sherr, B. F., E. B. Sherr, and R. D. Fallon. 1987. Use of monodispersed, fluorescently labeled bacteria to estimate in situ protozoan bacterivory. Appl. Environ. Microbiol. **53**: 958–965. doi:10.1016/0198-0254(87)90950-2

Sherr, B. F., E. B. Sherr, and J. McDaniel. 1992. Effect of protistan grazing on the frequency of

dividing cells in bacterioplankton assemblages. Appl. Environ. Microbiol. **58**: 2381–2385.

Sherr, E. B., and B. F. Sherr. 2002. Significance of predation by protists in aquatic microbial food webs. Antonie van Leeuwenhoek2 **81**: 293–308.

Sherr, E. B., B. F. Sherr, and P. A. Wheeler. 2005. Distribution of coccoid cyanobacteria and small eukaryotic phytoplankton in the upwelling ecosystem off the Oregon coast during 2001 and 2002. Deep. Res. Part II Top. Stud. Oceanogr. **52**: 317–330. doi:10.1016/j.dsr2.2004.09.020

Shi, X. L., D. Marie, L. Jardillier, D. J. Scanlan, and D. Vaulot. 2009. Groups without cultured representatives dominate eukaryotic picophytoplankton in the oligotrophic South East Pacific Ocean. PLoS One **4**. doi:10.1371/journal.pone.0007657

Shpak, M., J. R. Kugelman, A. Varela-Ramirez, and R. J. Aguilera. 2008. The phylogeny and evolution of deoxyribonuclease II: An enzyme essential for lysosomal DNA degradation. Mol. Phylogenet. Evol. **47**: 841–854. doi:10.1016/j.ympev.2007.11.033

Sieger, R., H. Grobe, M. Diepenbroek, U. Schindler, and R. Schlitzer. 2005. International Collection of JGOFS (Joint Global Ocean Flux Study), Volume 2: Integrated Data Sets (1989-2003), PANGAEA.

Silsbe, G. M., M. J. Behrenfeld, K. H. Halsey, A. J. Milligan, and T. K. Westberry. 2016. The CAFE model: A net production model for global ocean phytoplankton. Global Biogeochem. Cycles **30**: 1756–1777. doi:10.1002/2016GB005521

Simão, F. A., R. M. Waterhouse, P. Ioannidis, E. V. Kriventseva, and E. M. Zdobnov. 2015. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. Bioinformatics **31**: 3210–3212. doi:10.1093/bioinformatics/btv351

Six, C., Z. V Finkel, F. Rodriguez, D. Marie, F. Partensky, and D. A. Campbell. 2008. Contrasting photoacclimation costs in ecotypes of the marine eukaryotic picoplankter Ostreococcus. Limnol. Oceanogr. **53**: 255–265. doi:10.4319/lo.2008.53.1.0255

Smith, R. C., and K. S. Baker. 1982. Oceanic chlorophyll concentrations as determined by satellite (Nimbus-7 Coastal Zone Color Scanner). Mar. Biol. **66**: 269–279. doi:10.1007/BF00397032

Smythe-Wright, D., P. Chapman, C. Duncombe Rae, L. V. Shannon, and S. M. Boswell. 1998. Characteristics of the South Atlantic subtropical frontal zone between 15°W and 5°E. Deep. Res. Part I Oceanogr. Res. Pap. **45**: 167–192. doi:10.1016/S0967-0637(97)00068-X

Sosik, H. M., S. W. Chisholm, and R. J. Olson. 1989. Chlorophyll fluorescence from single cells : Interpretation of flow cytometric signals. Limnol. Oceanogr. **34**: 1749–1761.

Spalding, M. D., V. N. Agostini, J. Rice, and S. M. Grant. 2012. Pelagic provinces of the world: A biogeographic classification of the world's surface pelagic waters. Ocean Coast. Manag. **60**: 19–30. doi:10.1016/j.ocecoaman.2011.12.016

Steemann Nielsen, E. 1952. The use of radioactive carbon (C14) for measuring organic production in the sea. ICES J. Mar. Sci. **18**: 117–140.

Stenegren, M., A. Caputo, C. Berg, S. Bonnet, and R. A. Foster. 2017. Distribution and drivers of symbiotic and free-living diazotrophic cyanobacteria in the tropical South West Pacific. Biogeosciences Discuss. 1–47. doi:10.5194/bg-2017-63

Stramski, D., R. A. Reynolds, M. Babin, and others. 2008. Relationships between the surface concentration of particulate organic carbon and optical properties in the eastern South Pacific and eastern Atlantic Oceans. Biogeosciences **5**: 171–201. doi:10.5194/bg-5-171-2008

Stramski, D., A. Shalapyonok, and R. A. Reynolds. 1995. Optical characterization of the oceanic unicellular cyanobacterium Synechococcus grown under a day-night cycle in natural irradiance. J. Geophys. Res. **100**. doi:10.1029/95jc00452

Strickland, J. D. H., and T. R. Parsons. 1972. A Practical Handbook of Seawater Analyses, Fisheries Research Board of Canada.

Strzepek, R. F., P. W. Boyd, and W. G. Sunda. 2019. Photosynthetic adaptation to low iron, light, and temperature in Southern Ocean phytoplankton. Proc. Natl. Acad. Sci. U. S. A. **116**: 4388–4393. doi:10.1073/pnas.1810886116

Sunda, W. G., and S. A. Huntaman. 1997. Interrelated influence of iron, light and cell size on marine phytoplankton growth. Nature **390**: 389–392. doi:10.1038/37093

Suzuki, M. 1999. Effect of protistan bacterivory on coastal bacterioplankton diversity. Aquat. Microb. Ecol. **20**: 261–272. doi:10.3354/ame020261

Sverdrup, H. U. 1953. On Conditions for the Vernal Blooming of Phytoplankton. ICES J. Mar. Sci. **18**: 287–295. doi:10.1093/icesjms/18.3.287

Thorndsen, J. 1988. *Cymbomonas Schiller* (Prasinophyceae) reinvestigated by light and electron microscopy. Arch. fur Protistenkd. **136**: 327–336.

Thrash, J. C., J.-C. Cho, S. Ferriera, J. Johnson, K. L. Vergin, and S. J. Giovannoni. 2010. Genome sequences of *Pelagibaca bermudensis* HTCC2601T and *Maritimibacter alkaliphilus* HTCC2654T, the type strains of two marine Roseobacter genera. J. Bacteriol. **192**: 5552–5553. doi:10.1128/JB.00873-10

Tittel, J., V. Bissinger, B. Zippel, U. Gaedke, E. Bell, A. Lorke, and N. Kamjunke. 2003. Mixotrophs combine resource use to outcompete specialists: Implications for aquatic food webs. Proc. Natl. Acad. Sci. **100**: 12776–12781. doi:10.1073/pnas.2130696100

Turmel, M., A. Lopes dos Santos, C. Otis, R. Sergerie, and C. Lemieux. 2019. Tracing the evolution of the plastome and mitogenome in the Chloropicophyceae uncovered convergent tRNA gene losses and a variant plastid genetic code J. Archibald [ed.]. Genome Biol. Evol. **11**: 1275–1292. doi:10.1093/gbe/evz074

Uitz, J., H. Claustre, F. B. Griffiths, J. Ras, N. Garcia, and V. Sandroni. 2009. A phytoplankton class-specific primary production model applied to the Kerguelen Islands region (Southern Ocean). Deep. Res. Part I Oceanogr. Res. Pap. **56**: 541–560. doi:10.1016/j.dsr.2008.11.006

Uitz, J., H. Claustre, A. Morel, and S. B. Hooker. 2006. Vertical distribution of phytoplankton communities in open ocean: An assessment based on surface chlorophyll. J. Geophys. Res. Ocean. **111**. doi:10.1029/2005JC003207

Unrein, F., J. M. Gasol, F. Not, I. Forn, and R. Massana. 2014. Mixotrophic haptophytes are key bacterial grazers in oligotrophic coastal waters. ISME J. **8**: 164–176. doi:10.1038/ismej.2013.132

Vaqué, D., E. O. Casamayor, and J. M. Gasol. 2001. Dynamics of whole community bacterial production and grazing losses in seawater incubations as related to the changes in the proportions of bacteria with different DNA content. Aquat. Microb. Ecol. **25**: 163–177. doi:10.3354/ame025163

Vaulot, D., C. Courties, and F. Partensky. 1989. A Simple Method to Preserve Oceanic Phytoplankton for Flow Cytometric Analyses. Cytometry **10**: 629–635. doi:10.1002/cyto.990100519

Vazquez-Dominguez, E., F. Peters, J. M. Gasol, and D. Vaqué. 1999. Measuring the grazing losses of picoplankton: Methodological improvements in the use of fluorescently labeled tracers combined with flow cytometry. Aquat. Microb. Ecol. **20**: 119–128. doi:10.3354/ame020119

Veldhuis, M. J. W., and G. W. Kraay. 1993. Cell aubndance and fluorescence of picoplankton in relation to growth irradiance and nitrogen availability in the Red Sea. Netherland **31**: 135–145.

Veldhuis, M. J. W., and G. W. Kraay. 2000. Application of flow cytometry in marine phytoplankton research: current applications and future perspectives. Sci. Mar. **64**: 121–134. doi:10.3989/scimar.2000.64n2121

Veldhuis, M. J. W., G. W. Kraay, J. D. L. Van Bleijswijk, and M. A. Baars. 1997. Seasonal and spatial variability in phytoplankton biomass, productivity and growth in the northwestern Indian ocean: The southwest and northeast monsoon, 1992-1993. Deep. Res. Part I Oceanogr. Res. Pap. **44**: 425–449. doi:10.1016/S0967-0637(96)00116-1

Venables, H., and C. M. Moore. 2010. Phytoplankton and light limitation in the Southern Ocean: Learning from high-nutrient, high-chlorophyll areas. J. Geophys. Res. **115**. doi:10.1029/2009jc005361

Venter, C. J., K. Remington, J. F. Heidelberg, and others. 2004. Environmental genome shotgun sequencing of the Sargasso Sea. Science **304**: 66–74. doi:10.11 26/science.1093857

de Verneil, A., L. Rousselet, A. M. Doglioli, A. A. Petrenko, C. Maes, P. Bouruet-Aubertot, and T. Moutin. 2018. OUTPACE long duration stations: physical variability, context of

biogeochemical sampling, and evaluation of sampling strategy. Biogeosciences **15**: 2125–2147. doi:10.5194/bg-15-2125-2018

de Verneil, A., L. Rousselet, A. M. Doglioli, A. A. Petrenko, and T. Moutin. 2017. The fate of a southwest Pacific bloom: gauging the impact of submesoscale vs. mesoscale circulation on biological gradients in the subtropics. Biogeosciences **14**: 3471–3486. doi:10.5194/bg-14-3471-2017

Vidussi, F., H. Claustre, B. B. Manca, A. Luchetta, and M. Jean-Claude. 2001. Phytoplankton pigment distribution in relation to upper thermocline circulation in the eastern Mediterranean Sea during winter. J. Geophys. Res. **106**: 939–956.

Walter, B., J. Peters, J. E. E. van Beusekom, and M. A. St. John. 2015. Interactive effects of temperature and light during deep convection: a case study on growth and condition of the diatom Thalassiosira weissflogii. ICES J. Mar. Sci. **72**: 2061–2071. doi:10.1093/icesjms/fsu218

Van Wambeke, F., P. Catala, M. Pujo-Pay, and P. Lebaron. 2011. Vertical and longitudinal gradients in HNA-LNA cell abundances and cytometric characteristics in the Mediterranean Sea. Biogeosciences **8**: 1853–1863. doi:10.5194/bg-8-1853-2011

Van Wambeke, F., A. Gimenez, S. Duhamel, C. Dupouy, D. Lefevre, M. Pujo-Pay, and T. Moutin. 2018. Dynamics of phytoplankton and heterotrophic bacterioplankton in the western tropical South Pacific Ocean along a gradient of diversity and activity of diazotrophs. Biogeosciences Discuss. 1–32. doi:10.5194/bg-2017-556

Ward, B. A., and M. J. Follows. 2016. Marine mixotrophy increases trophic transfer efficiency, mean organism size, and vertical carbon flux. Proc. Natl. Acad. Sci. U. S. A. **113**: 2958–2963. doi:10.1073/pnas.1517118113

Ward, B. B., A. H. Devol, J. J. Rich, B. X. Chang, S. E. Bulow, H. Naik, A. Pratihary, and A. Jayakumar. 2009. Denitrification as the dominant nitrogen loss process in the Arabian Sea. Nature **461**: 78–81. doi:10.1038/nature08276

Weisse, T., R. Anderson, H. Arndt, A. Calbet, P. J. Hansen, and D. J. S. Montagnes. 2016. Functional ecology of aquatic phagotrophic protists – Concepts, limitations, and perspectives. Eur. J. Protistol. **55**: 50–74. doi:10.1016/j.ejop.2016.03.003

Welschmeyer, N. A. 1994. Fluorometric analysis of chlorophyll a in the presence of chlorophyll b and pheopigments. Limnol. Oceanogr. **39**: 1985–1992.

Westberry, T., M. J. Behrenfeld, D. A. Siegel, and E. Boss. 2008. Carbon-based primary productivity modeling with vertically resolved photoacclimation. **22**: 1–18. doi:10.1029/2007GB003078

Whitmire, A., R. Ltelier, V. Villagrán, and O. Ulloa. 2009. Autonomous observations of in vivo fluorescence and particle backscattering in an oceanic oxygen minimum zone. Opt. Express **17**.

Wickham, H. 2009. ggplot2: Elegant Graphics for Data Analysis, Springer-Verlag.

Wiggert, J. D., R. R. Hood, K. Banse, and J. C. Kindle. 2005. Monsoon-driven biogeochemical processes in the Arabian Sea. Prog. Oceanogr. **65**: 176–213. doi:10.1016/j.pocean.2005.03.008

Wilken, S., C. C. M. Yung, M. Hamilton, and others. 2019. The need to account for cell biology in characterizing predatory mixotrophs in aquatic environments. Philos. Trans. R. Soc. B Biol. Sci. **374**. doi:10.1098/rstb.2019.0090

Wilson, C., and V. J. Coles. 2005. Global climatological relationships between satellite biological and physical observations and upper ocean properties. J. Geophys. Res. C Ocean. **110**: 1–14. doi:10.1029/2004JC002724

Wincker, P. 2019. A thousand plants' phylogeny. Nat. Plants **5**: 1106–1107. doi:10.1038/s41477-019-0555-0

Wong, A. P., R. Keeley, and T. Carval. 2020. Argo Quality Control Manual For CTD and Trajectory Data.

Worden, A. 2006. Picoeukaryote diversity in coastal waters of the Pacific Ocean. Aquat. Microb. Ecol. **43**: 165–175. doi:10.3354/ame043165

Wright, S., S. Jeffrey, and R. Mantoura. 1997. Evaluation of methods and solvents for pigment extraction, p. 261–282. *In* S. Jeffrey, R. Mantoura, and S. Wright [eds.], Phytoplankton pigments in oceanography: guidelines to modern methods. UNESCO.

Yabuki, A., Y. Inagaki, and K. Ishida. 2010. *Palpitomonas bilix* gen. et sp. nov.: A Novel deep-branching heterotroph possibly related to Archaeplastida or Hacrobia. Protist **161**: 523–538. doi:10.1016/j.protis.2010.03.001

Yentsch, C. M., P. K. Horan, K. Muirhead, and others. 1983. Flow cytometry and cell sorting: A technique for analysis and sorting of aquatic particles. Limnol. Oceanogr. **28**: 1275–1280. doi:10.4319/lo.1983.28.6.1275

Yoo, Y. Du, K. A. Seong, H. J. Jeong, W. Yih, J.-R. Rho, S. W. Nam, and H. S. Kim. 2017. Mixotrophy in the marine red-tide cryptophyte *Teleaulax amphioxeia* and ingestion and grazing impact of cryptophytes on natural populations of bacteria in Korean coastal waters. Harmful Algae **68**: 105–117. doi:10.1016/j.hal.2017.07.012

Zingone, A., M. Borra, C. Brunet, G. Forlani, W. H. C. F. Kooistra, and G. Procaccini. 2002. Phylogenetic position of *Crustomatix stigmatica* sp. nov. and *Dolichomastix tenuilepis* in relation to the mamiellales (Prasinophyceae, Chlorophyta). J. Phycol. **38**: 1024–1039. doi:10.1046/j.1529-8817.2002.t01-1-02035.x

Zubkov, M. V. 2014. Faster growth of the major prokaryotic versus eukaryotic CO2 fixers in the oligotrophic ocean. Nat. Commun. **5**: 1–6. doi:10.1038/ncomms4776

Zubkov, M. V., P. H. Burkill, and J. N. Topping. 2007. Flow cytometric enumeration of DNA-stained oceanic planktonic protists. J. Plankton Res. **29**: 79–86. doi:10.1093/plankt/fbl059

Zubkov, M. V., M. A. Sleigh, P. H. Burkill, and R. J. G. Leakey. 2000. Picoplankton community structure on the Atlantic Meridional Transect: A comparison between seasons. Prog. Oceanogr. **45**: 369–386. doi:10.1016/S0079-6611(00)00008-2

Zubkov, M. V., and G. A. Tarran. 2008. High bacterivory by the smallest phytoplankton in the North Atlantic Ocean. Nature **455**: 224–226. doi:10.1038/nature07236

Appendix A: Chapter 2 Supplementary material

A.1. Derivation of SVD coefficients

EOF is a specific application of SVD: a matrix factorization of the form M = UΣV, where U is a matrix of left singular vectors, Σ is a matrix of singular values, and V is a matrix of right singular vectors. Each column in the input matrix M can be described as a linear combination of the columns of ΣV, based on coefficients in each column of matrix U, providing empirical orthogonal functions of the following form:

$$M_i = \sum_{k=1}^{P} U_{ki}(\Sigma V)_k$$

Because the columns of ΣV are constants, this makes it possible to describe the variability associated with each column of M based on the values of U alone. Moreover, because each column in ΣV captures a decreasing fraction of the total variance in the input matrix M, it is also possible to identify a subset of the right singular vectors in matrix ΣV that capture most of the variance in the original data. This way, it is possible to summarize the variability in each column of matrix M with only a small number of weightings from matrix U (Menke and Menke 2016).

A.2. Determination of clusters based on SVD coefficients

The number of weightings and the distance metric used in the cluster analysis were selected to maximize the average silhouette score of each cluster, calculated as

$$a(i) = \frac{\sum_{i \neq j}^{j \in C} d(i,j)}{|C_i| - 1}$$

Where *a* is the average distance between a data point *i* and the other *j* data points within its corresponding cluster, calculated as the distance *d* and normalized to 1 subtracted from the number of data points in the cluster. Data points with negative silhouette scores (implying that

they were on average closer to a cluster other than the cluster they were assigned to), were reassigned to their nearest neighboring cluster.

The clustering analysis was complicated by 1) subjectivity in the number of weightings to be included as clustering criteria and 2) by differences in the proportion of variability captured by each weighting for individual time series. For example, in the high-latitude floats, most variability occurred in weightings attributable to the first and second singular vectors. In the low-latitude floats variability was distributed across weightings corresponding to the first four singular vectors, and in the case of some Indian Ocean floats, across the first six singular vectors (Table A.1).

To identify the number of clusters and criteria that maximized the similarity of time series within each cluster, different combinations were tested iteratively for the high-latitude and low-latitude floats. This analysis demonstrated that the silhouette score was maximized for the high-latitude floats when forming two clusters based on the weightings corresponding to the first two singular vectors. The silhouette score was maximized for the low-latitude floats when forming four clusters based weightings corresponding to the first six singular vectors. Although a slightly higher silhouette score is achieved for low-latitude floats when forming four clusters based on the first two singular vectors, this improvement was caused by the formation

The silhouette score was maximized for the global clustering when using the Manhattan distance to compare the similarity of the first two weightings. The silhouette score was maximized for the clustering of the low-latitude floats when using the Manhattan distance to compare the similarity of the first two weightings. The silhouette score was maximized for the clustering of the high-latitude floats when using the Euclidean distance to compare the similarity of the first six weightings. The differences in the number of weightings required in order to

maximize the silhouette score is due to the weightings 1 & 2 containing the majority of the variability in the high-latitude floats, and weightings 4 – 6 containing most of the variability differentiating the low-latitude floats (Table A.2).

Table A.1

Round 1 clusters (N)	Round 1 Average silhouette score	Round 2 clusters (N)	Round 2 Average silhouette score
Low-latitude (187)	0.62	**EQ (59)**	0.48
		AR (18)	0.45
		SDCM (51)	0.42
		PDCM (59)	0.59
High-latitude (235)	0.62	**LCB (163)**	0.59
		HCB (72)	0.60

Table A.2

Biome	V_1	V_2	V_3	V_4	V_5	V_6	total
AR	59 ± 12	5.5 ± 4.3	9.2 ± 5.4	0.7 ± 1.0	1.7 ± 1.7	2.0 ± 1.6	78 ± 8.9
EQ	44 ± 9.6	0.9 ± 1.3	15 ± 4.7	6.1 ± 4.5	1.0 ± 1.0	2.9 ± 2.3	70 ± 8.6
SDCM	55 ± 14	17 ± 10	3.0 ± 3.3	1.2 ± 1.2	2.8 ± 2.8	1.3 ± 1.6	79 ± 8.1
PDCM	40 ± 9.8	27 ± 3.4	16 ± 8.0	3.1 ± 2.0	0.4 ± 0.4	0.3 ± 0.4	86 ± 5.7
LCB	74 ± 10	3.3 ± 2.5	1.8 ± 2.0	2.8 ± 3.1	2.0 ± 2.6	0.9 ± 1.4	84 ± 8.0
HCB	50 ± 11	18 ± 6.1	2.0 ± 1.8	2.8 ± 2.8	1.6 ± 2.5	1.5 ± 2.4	76 ± 10

Table A.4

Results for linear regressions comparing log transformed monthly averages of bbp_{mz} to monthly averages of $iPAR_{NO3}$ and N_{MLD} for each cluster. Log transformations were performed to linearize relationships.

	$iPAR_{NO3}$			N_{MLD}		
	slope	**p-value**	**R^2**	**slope**	**p-value**	**R^2**
EQ	0.05	0.08	0.28	-0.06	0.01	0.50
AR	-0.04	0.03	0.37	0.14	0.1	0.24
PDCM	0.03	0.30	0.10	-0.07	0.23	0.14
SDCM	0.004	0.91	0.01	0.03	0.60	0.02
LCB	0.15	3×10^{-7}	0.94	-4.9	0.01	0.47
HCB	0.20	2×10^{-6}	0.91	-1.1	0.001	0.65

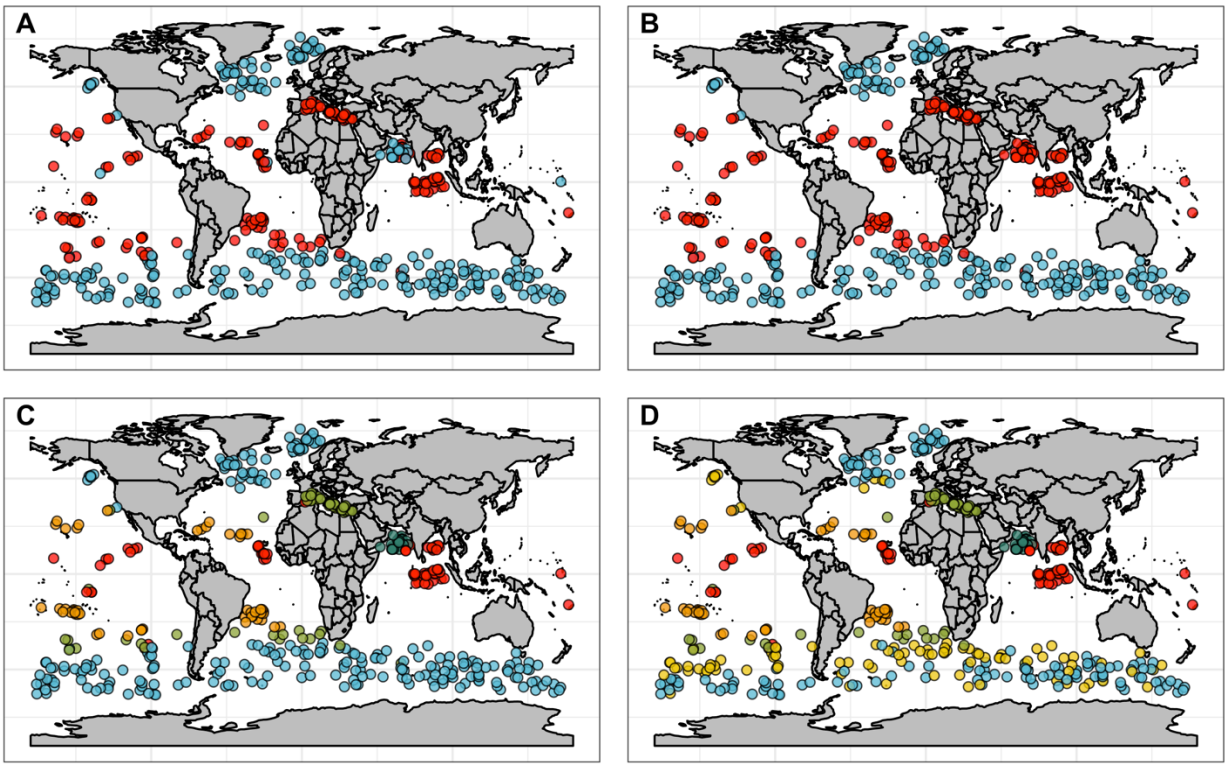

Figure A.1

Illustration of the results achieved by clustering procedure used to obtain the 6 identified clusters. A) Results of global cluster analysis based on Euclidean distance of first 3 vectors. B) Results of global cluster analysis after reassignment of misclassified low-latitude floats; C) Global distribution of clusters after incorporating cluster results for low-latitude floats, based on Euclidean distance of first 6 vectors; D) Global distribution of clusters after incorporating cluster results for high-latitude floats, based on Manhattan distance of first 3 vectors. The clustering method and number of vectors used at each step was selected to maximize the average silhouette score of the resultant clusters.

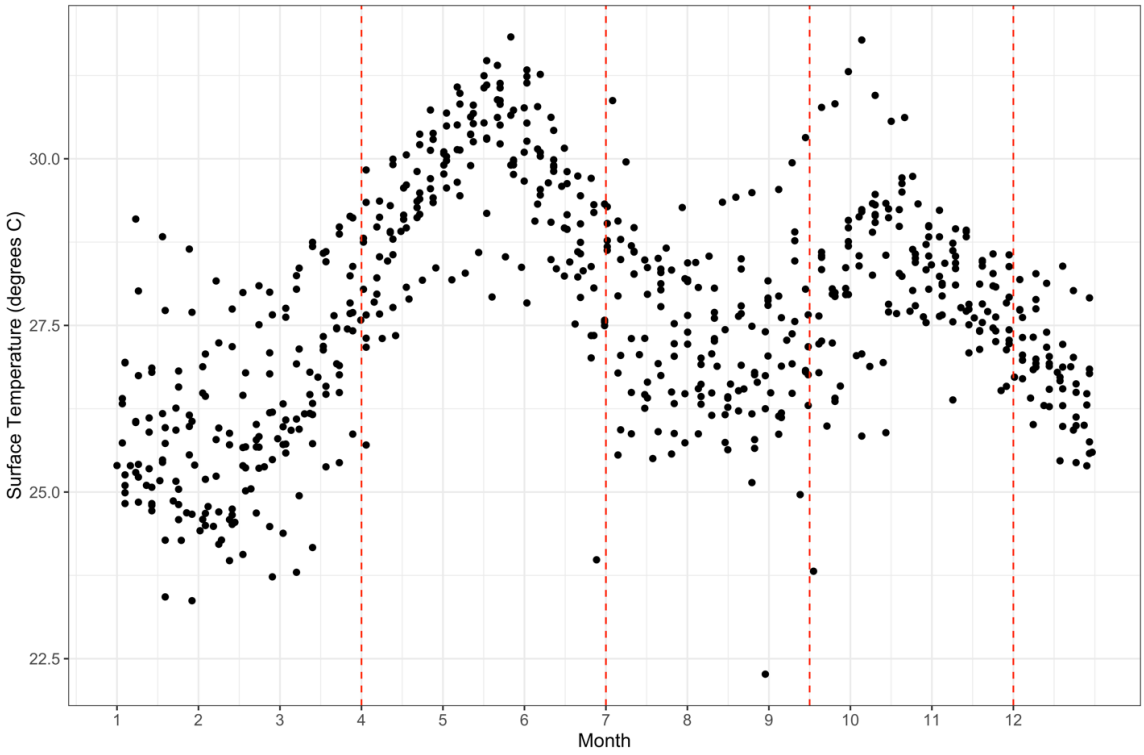

Figure A.2

Sea surface temperature values for profiles in AR time series. From left to right, red dashed lines correspond to start of spring intermonsoon (SIM), southwest monsoon (SWM), fall intermonsoon (FIM), and northeast monsoon (NEM).

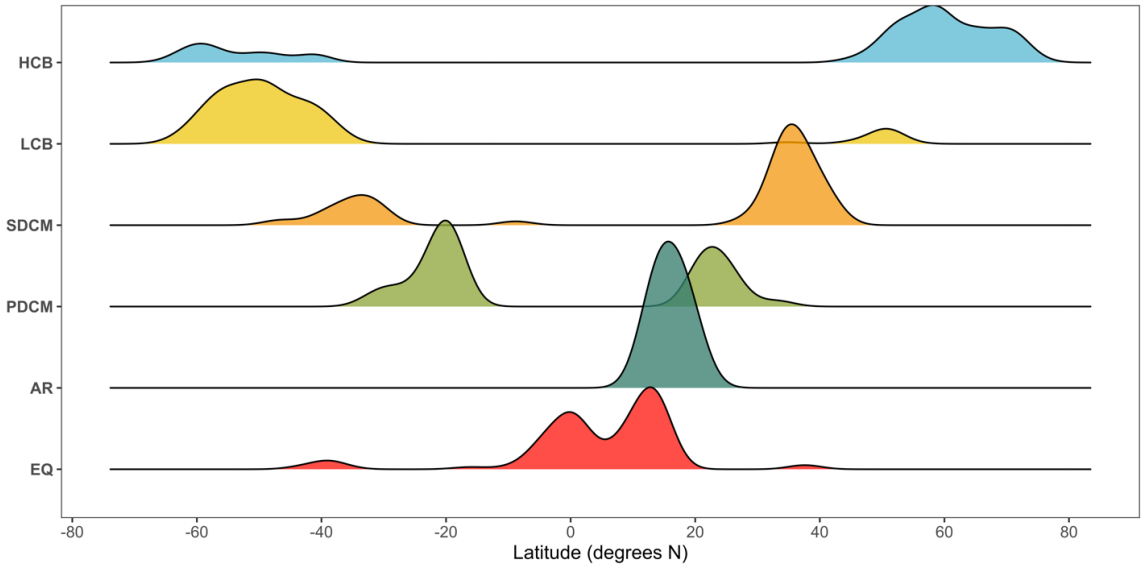

Figure A.3

Latitudinal distribution of profiles belonging to Equatorial (EQ), Arabian Sea (AR), Seasonal DCM (SDCM), Permanent DCM (PDCM), Low Chlorophyll Bloom (LCB), and High Chlorophyll Bloom (HCB) clusters. Height of ridge for each biome corresponds to the number of profiles at a given latitude.

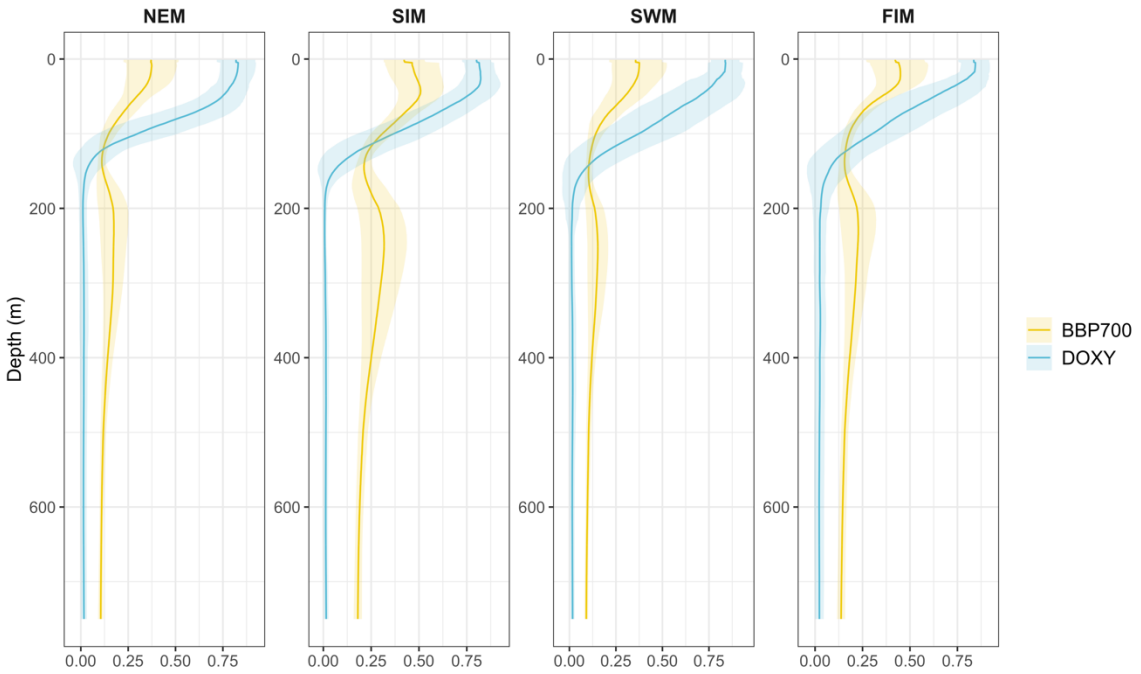

Figure A.4

Depth profiles of b_{bp} and dissolved oxygen (DOXY) for profiles in Arabian Sea (AR) biome. Panels correspond to Northeast Monsoon (NEM), Summer Intermonsoon (SIM), Southwest Monsoon (SWM) and Fall Intermonsoon (FIM). Each parameter is normalized to its maximum value for ease of comparison.

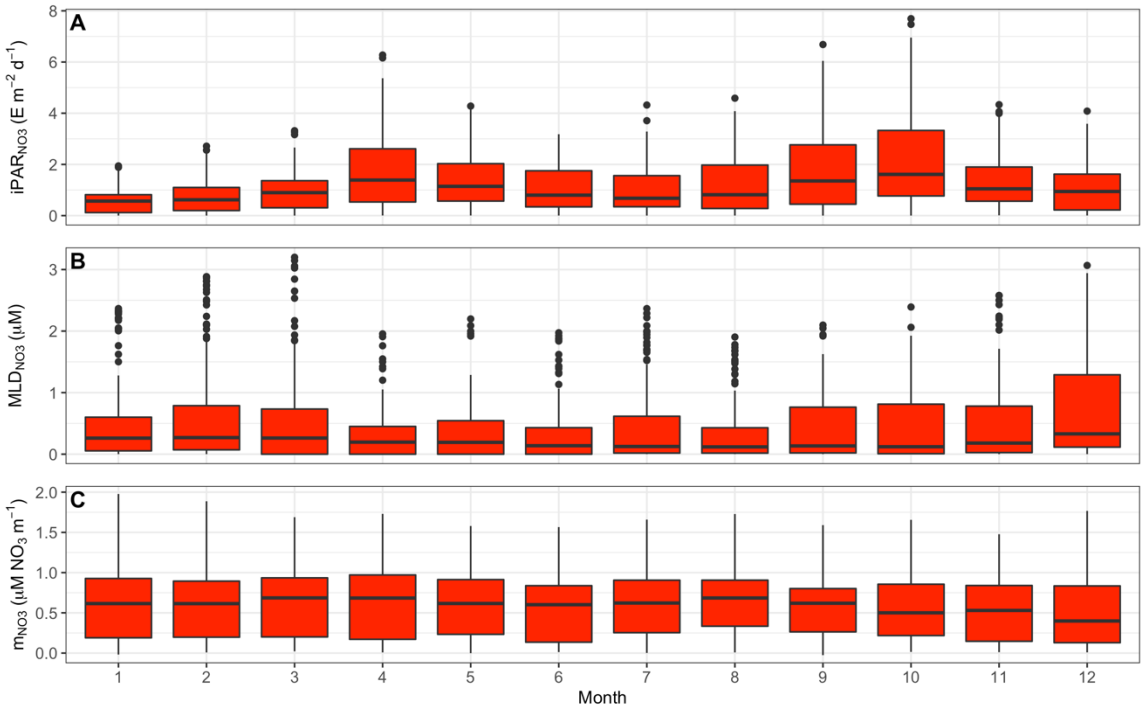

Figure A.5

Monthly distributions of data for **A)** [Chl*a*]$_{sat}$, **B)** iPAR$_{NO3}$, **C)** N$_{MLD}$, and **D)** m$_{NO3}$ for time series in Equatorial (EQ) biome. Outliers, defined as values greater than q_3 + (1.5*IQR) or less than q_1 - (1.5*IQR), have been removed to emphasize seasonal trends.

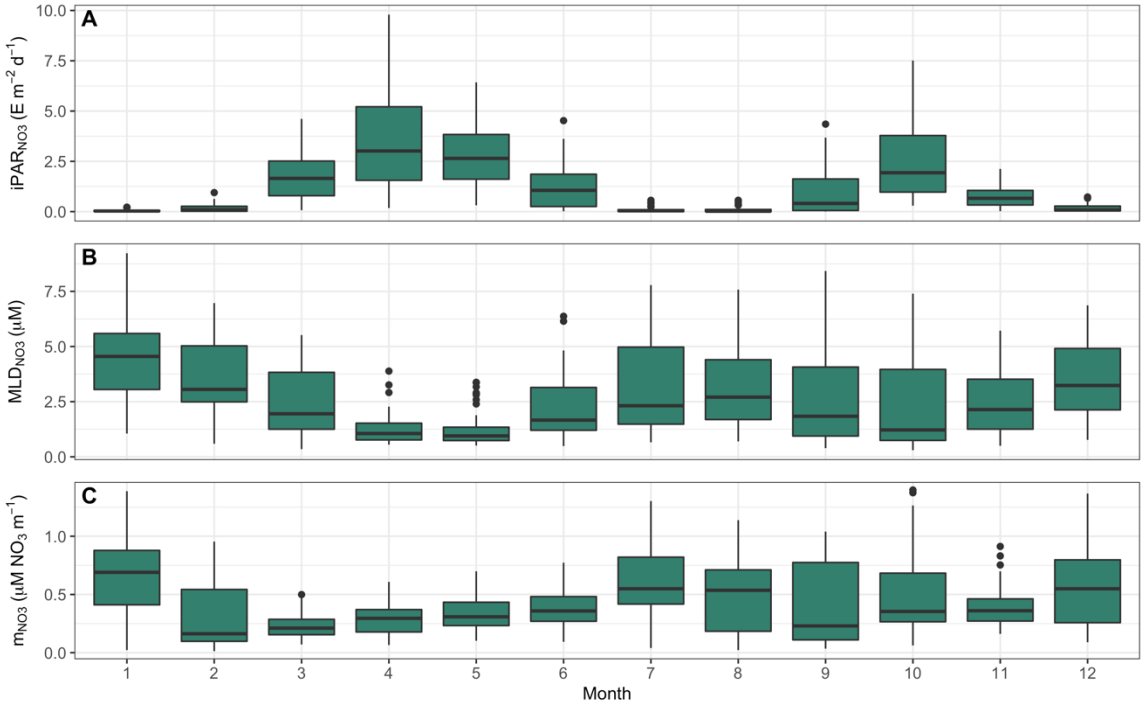

Figure A.6

Monthly distributions of data for **A)** $iPAR_{NO3}$, **B)** MLD_{NO3}, and **C)** m_{NO3} for time series in Arabian Sea (AR) biome. Outliers, defined as values greater than q + (1.5*IQR) or less than q_1 - (1.5*IQR), have been removed to emphasize seasonal trends.

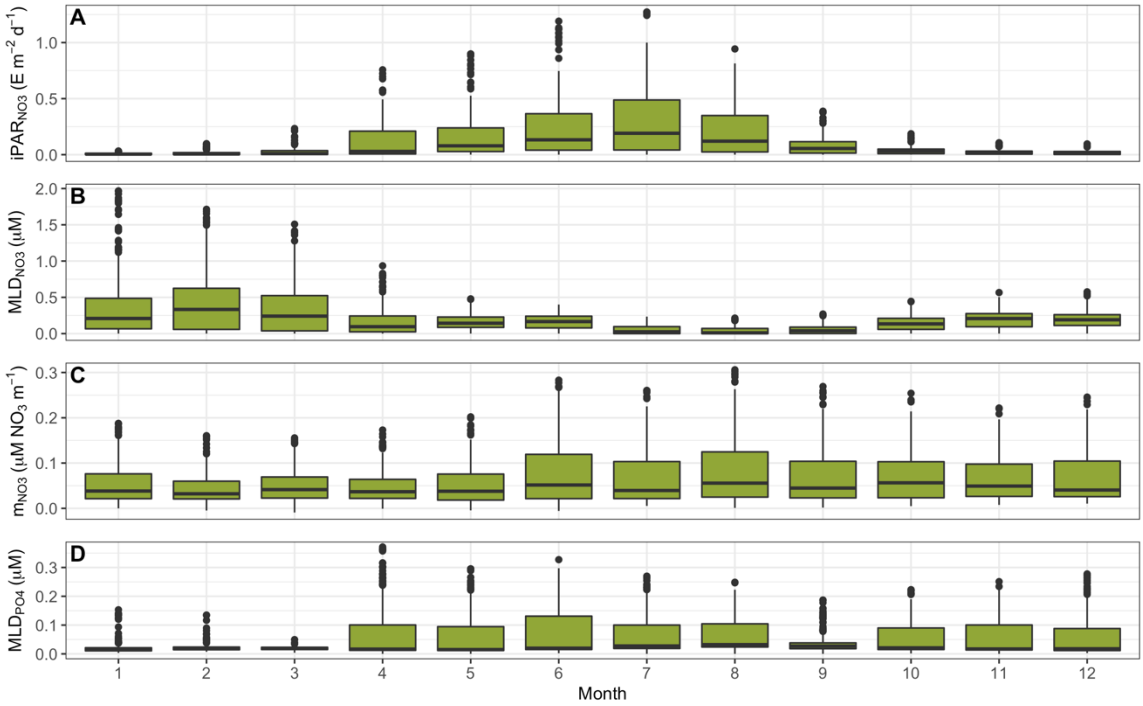

Figure A.7

Monthly distributions of data for **A)** $iPAR_{NO3}$, **B)** MLD_{NO3}, **C)** m_{NO3}, and **D)** MLD_{PO4} for time series in Seasonal DCM (SDCM) biome. Outliers, defined as values greater than q_3 + (1.5*IQR) or less than q_1 - (1.5*IQR), have been removed to emphasize seasonal trends.

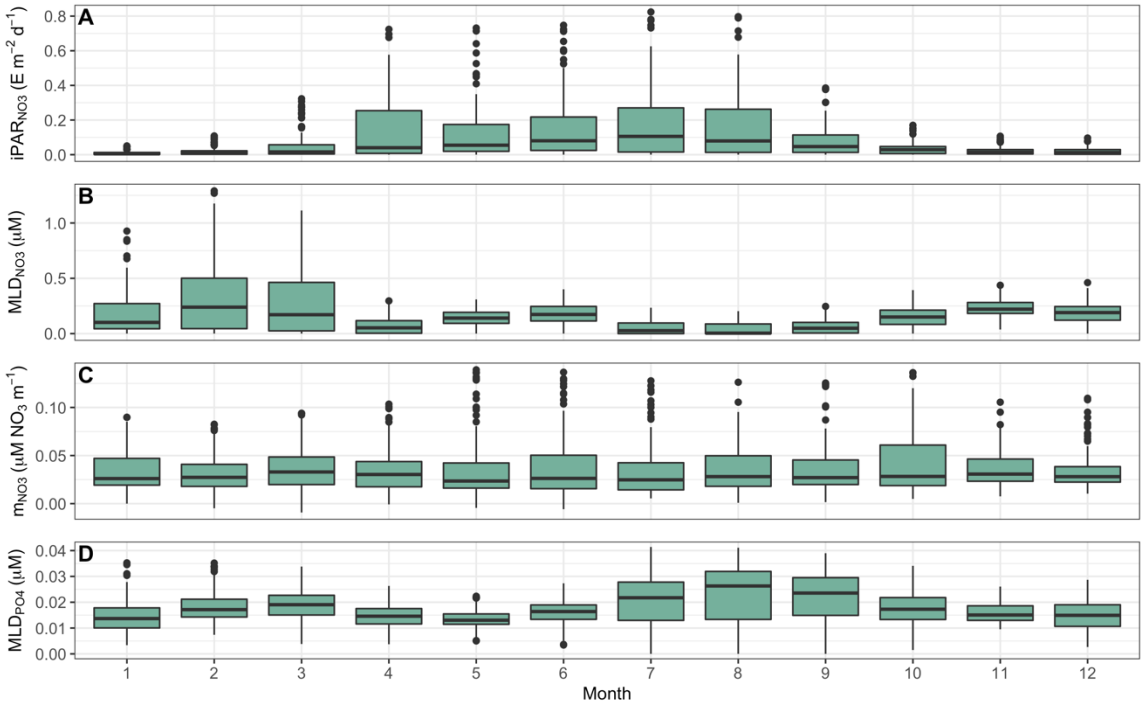

Figure A.8

Monthly distributions of data for **A)** $iPAR_{NO3}$, **B)** MLD_{NO3}, **C)** m_{NO3}, and **D)** MLD_{PO4} for Mediterranean (MED) time series. Outliers, defined as values greater than q_3 + (1.5*IQR) or less than q_1 - (1.5*IQR), have been removed to emphasize seasonal trends.

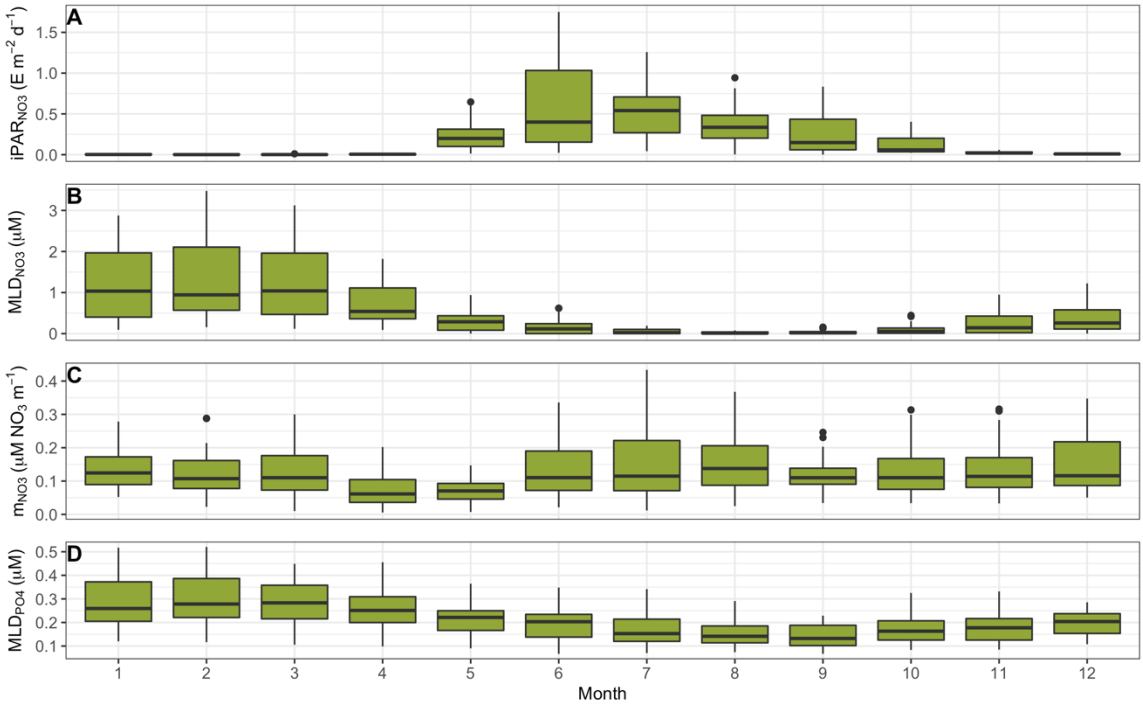

Figure A.9

Monthly distributions of data for **A)** $iPAR_{NO3}$, **B)** MLD_{NO3}, **C)** m_{NO3}, and **D)** MLD_{PO4} for Subtropical Frontal Zone (STFZ) time series. Outliers, defined as values greater than q_3 + (1.5*IQR) or less than q_1 - (1.5*IQR), have been removed to emphasize seasonal trends.

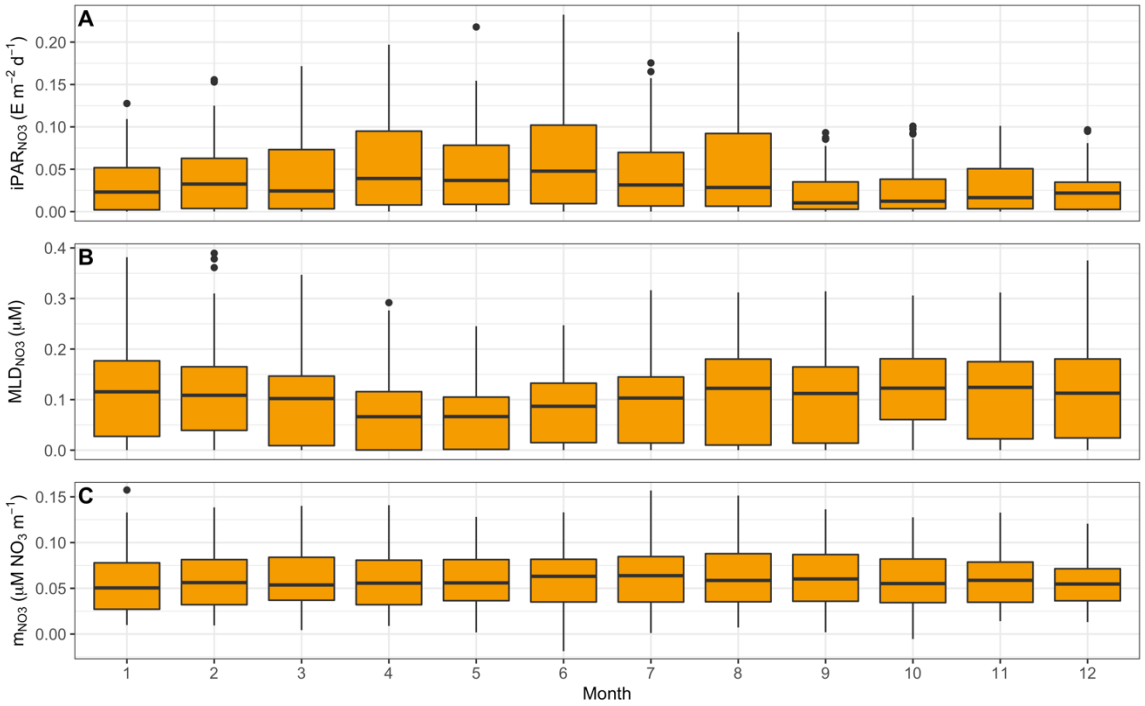

Figure A.10

Monthly distributions of data for **A)** $iPAR_{NO3}$, **B)** MLD_{NO3}, and **C)** m_{NO3} for time series in Permanent DCM (PDCM) biome. Outliers, defined as values greater than q_3 + (1.5*IQR) or less than q_1 - (1.5*IQR), have been removed to emphasize seasonal trends.

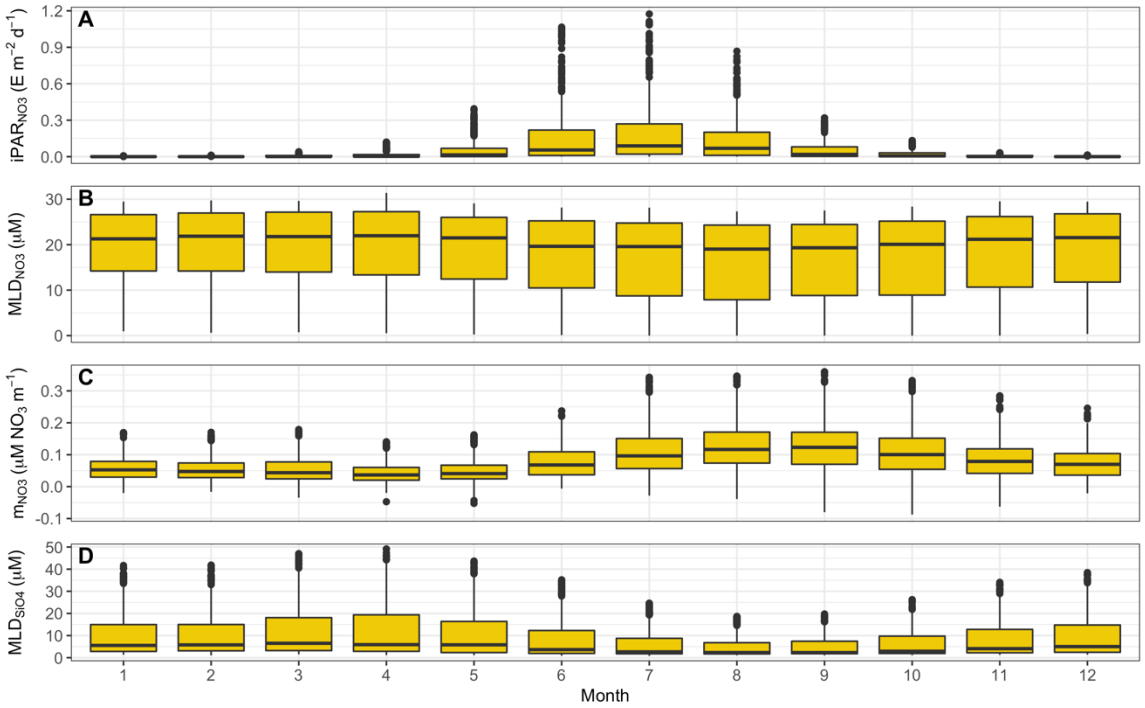

Figure A.11

Monthly distributions of data for **A)** $iPAR_{NO3}$, **B)** MLD_{NO3}, **C)** m_{NO3}, and **D)** MLD_{SiO4} for time series in Low Chlorophyll Bloom (LCB) biome. Outliers, defined as values greater than q_3 + (1.5*IQR) or less than q_1 - (1.5*IQR), have been removed to emphasize seasonal trends.

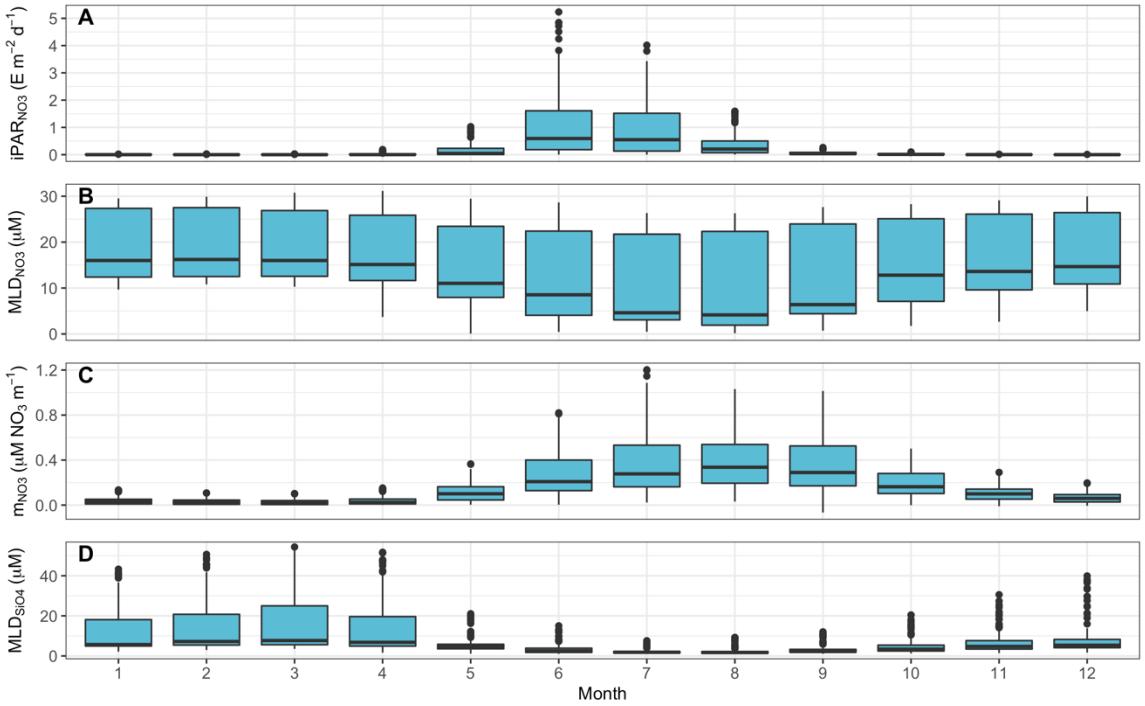

Figure A.12

Monthly distributions of data for **A)** $iPAR_{NO3}$, **B)** MLD_{NO3}, **C)** m_{NO3}, and **D)** MLD_{SiO4} for time series in High Chlorophyll Bloom (HCB) biome. Outliers, defined as values greater than q_3 + (1.5*IQR) or less than q_1 - (1.5*IQR), have been removed to emphasize seasonal trends.

Appendix B: Chapter 3 Supplementary material

B.1. Evaluation of narrow bandpass and broad band fluorometer filter sets

To compare quantitation and detection limits for the narrow bandpass (excitation: 436/10 nm, emission: 685/10 nm) and broadband filter sets (excitation ≤ 500 nm, emission ≥ 665 nm), fluorescence of triplicate Chl*a* calibration standards was measured using the broadband filter set. The narrow bandpass filter set produced lower variability in RFU values, despite lower overall signal intensity than the broadband filter set (Figure B.1), especially at low Chl*a* concentrations. The broadband filter set performed comparatively poorly at Chl*a* concentrations below 2×10^{-4} μg Chl, yielding regressions with R^2 values of 0.75, compared to R^2 values of 0.99 in measurements taken with the narrow bandpass filter set. Similarly, the relative error of measurements of methanol blanks was 3.7 % for the narrow bandpass filter set and 4.9 % for the broadband filter set. Detection and quantitation limits using the narrow bandpass filter set were 3.21×10^{-6} and 1.07×10^{-5} μg Chl*a*, while those for broadband filter sets were 1.02×10^{-4} and 3.41×10^{-4} μg Chl*a*, respectively.

B.2. Results of filter set evaluation

We found narrow bandpass filters to considerably lower the number of cells per filter required to exceed fluorometer quantitation limits. For example, at the sorting thresholds used, 40 % of samples were below the quantitation limit of the broadband filter set. The decreased detection and quantitation limits encountered when using the narrow bandpass filter sets were primarily a result of the increased sensitivity of the narrow bandpass filters compared to the broadband filters. The measurement response per unit increase in Chl*a*, relative to blank measurements, was approximately 10 times greater in the narrow bandpass filter set compared to the broadband filter

set, meaning that a given multiple of a blank measurement corresponded to a lower Chl*a* concentration for the narrow bandpass filters compared to the broadband filters (Figure B.1).

B.3. Test of variability resulting from sample storage duration

To assess the effect of sample preparation and storage duration on per-cell Chl*a*, cell suspensions of cultured *Prochlorococcus* CCMP 2777, *Synechococcus* WH2777, and *M. squamata* NIES 1407 were prepared in artificial seawater to the same final abundance as used for the extraction duration tests (previous section). One-ml subsamples of this cell suspension were transferred to cryovials, flash frozen in liquid nitrogen without the addition of biological fixatives and stored at -80° C until processing. One set of triplicate samples was sorted after 1 hour storage at -80° C to determine initial per-cell Chl*a* content. Additional triplicate samples were thawed after 2 month, 3 month, and 8 month intervals and then sorted as described above. To ensure sufficient biomass to exceed fluorometer detection limits across all time points, final concentrations of 5×10^5 cells per filter for *Prochlorococcus* and *Synechococcus*, and 1.5×10^4 cells per filter for *M. squamata* were sorted at each time point. Chl*a* extractions were then performed as described above. Flow cytometric cell counts of *Prochlorococcus*, *Synechococcus*, and PPE were also performed for each thawed sample at each time point.

B.4. Results of storage duration tests

Average initial per-cell Chl*a* values for storage duration tests were 0.78 ± 0.03 fg Chl*a* cell^{-1} for *Prochlorococcus*, 1.35 ± 0.03 fg Chl*a* cell^{-1} for *Synechococcus*, and 10.01 ± 0.30 fg Chl*a* cell^{-1} for PPE. There were no statistically significant differences in average per-cell Chl*a* in samples collected over 8 months of storage for any phytoplankton groups (ANOVA; $p > 0.05$ for all comparisons). Average initial cell abundance for storage duration tests were $3.14 \pm 0.24 \times 10^5$ for *Prochlorococcus*, $2.26 \pm 0.16 \times 10^5$ for *Synechococcus*, and $0.13 \pm 0.01 \times 10^5$ for PPE. Cell

abundance decreased by 62.2 ± 12.5 %, 17.3 ± 9.3 %, and 81.3 ± 3.8 % between 0 and 8 month samples for *Prochlorococcus*, *Synechococcus*, and PPE, respectively (Figure 3.3). These decreases were statistically significant for *Prochlorococcus* and PPE (ANOVA; $p > 0.05$ for all comparisons).

B.5. Test of variability resulting from extraction duration

To evaluate the effect of methanol extraction duration on measured Chl*a* concentration, methanol extractions were performed on triplicate samples taken from laboratory cultures of *Prochlorococcus* CCMP 2777, *Synechococcus* WH 8102, and the pico-sized chlorophyte *Mantoniella squamata* NIES 1407. Cultures were diluted separately to final densities of 1×10^6 cells ml^{-1} for *Prochlorococcus* and *Synechococcus*, and 1×10^5 cells ml^{-1} for *M. squamata.* One-ml aliquots from each dilution were filtered onto 25-mm diameter 0.2-µm pore polycarbonate filters (Sigma Millipore catalog #GTTP02500) that were then placed in cryovials and frozen at -80° C for 1 hour. Samples were thawed in the dark to room temperature and placed in 15-ml plastic centrifuge tubes (Falcon®) with 10 ml 100% methanol. Samples were vortexed for 15 seconds, and then stored in the dark at -4° C. Chl*a* was then measured as described above after extracting for 1, 14, 18, 22, 40, and 44 hours.

B.6. Results of extraction duration tests

Methanol extracts achieved maximal fluorescence after 1 hour of incubation for all strains (data not shown). There were no statistically significant differences between average RFU values when comparing measurements made after 1 hour to those from longer extraction durations (ANOVA; $p > 0.05$ for all groups).

Table B.1

Summary of variability in measurements of per-cell Chl*a* for *Prochlorococcus* (PRO), *Synechococcus* (SYN), and picophytoeukaryotes (PPE). Per-cell Chl*a* for each group calculated based on filter Chl*a* content normalized to the number of cells on each filter (NORM) or based on regressions between filter Chl*a* content and the number of cells on each filter (REG).

Group	n	Calculation method	fg Chl*a* cell^{-1}	Relative error	Chl_{reg} - Chl_{norm} (fg Chla cell^{-1})
PRO	14	NORM	0.50 ± 0.009	2.05 ± 0.45 %	-0.002 ± 0.004
		REG	0.50 ± 0.012	2.87 ± 0.51 %	
SYN	19	NORM	0.92 ± 0.017	2.05 ± 0.32 %	-0.001 ± 0.005
		REG	0.92 ± 0.023	2.76 ± 0.34 %	
PPE	16	NORM	37.68 ± 0.092	2.70 ± 0.28 %	-0.41 ± 0.31
		REG	37.28 ± 1.43	4.19 ± 0.38 %	

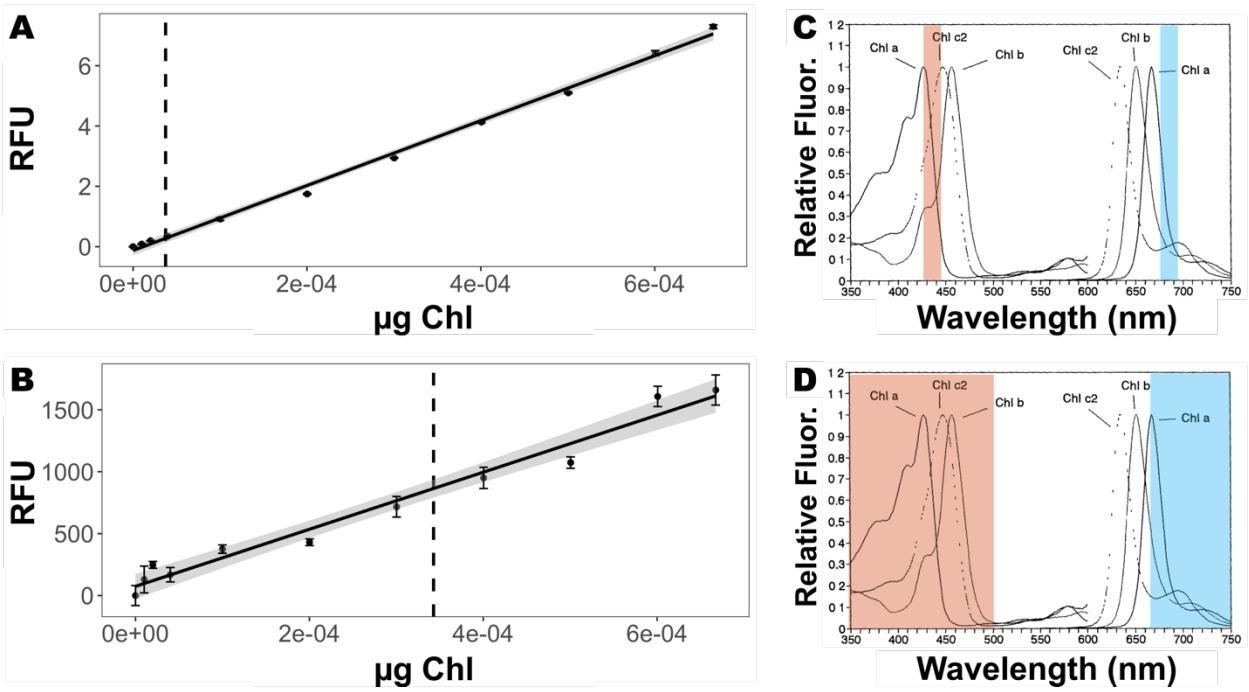

Figure B.1: Calibration curves prepared from 1 x 10^{-5} to 6.7 x 10^{-4} µg Chl*a* using the narrow bandpass **(A)** and the broadband filter sets **(B)**. Error bars correspond to standard error of measurements at each concentration. Grey shading shows ± 1 standard error of regression coefficients. Vertical dashed lines show instrument quantitation limits. All measurements were zeroed on a methanol blank. **(B)** and **(D)**: Excitation and emission bands of narrow bandpass and broadband filter sets, respectively, superimposed on chlorophyll fluorescence spectra (from Welshmeyer et al. (1994)).

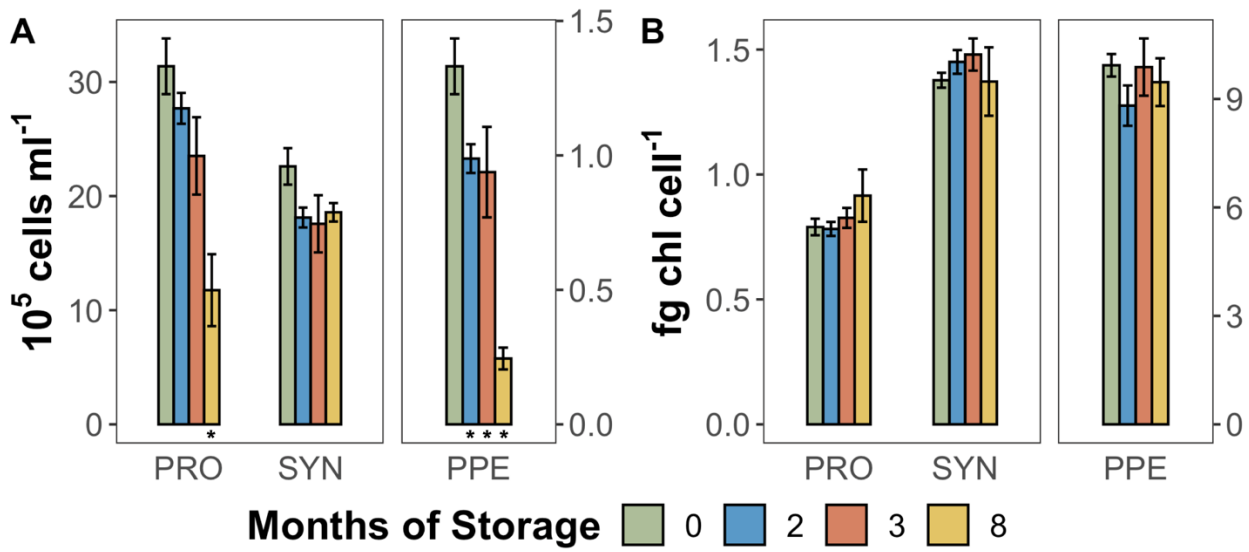

Figure B.2: **(A)** Cell abundance in samples counted flow cytometrically after 0, 2, 3, or 8 months storage at -80° C for *Prochlorococcus* (PRO), *Synechococcus* (SYN), and *M. squamata* (PPE). **B)** Per-cell Chl*a* (fg Chl cell^{-1}) in sorted and filtered samples extracted after 0, 2, 3, and 8 months storage at -80° C. Error bars correspond to ± 1 standard error of triplicate sample means. Asterisks indicate time-points where average values were significantly different than those corresponding to 0 month samples based on ANOVA followed by Tukey's Honestly Significant Difference post-hoc test.

Appendix C: Chapter 5 Supplementary material

C.1. Prey viability tests method

To test the effect of the Cell Tracker CMFDA labeling protocol on the viability of *Pelagibaca bermudensis*, three dilutions of the bacterial culture were prepared at late exponential phase to final cell abundances of ~ 1 x 10^9 cells mL^{-1} in 1 mL of ASW. One dilution was maintained at normal growth temperatures of 25° C; one dilution was incubated for three hours at 37° C, washed three times in ASW, then stored at 4° C; one dilution was inoculated with CT, incubated for three hours at 37° C, washed three times in ASW, then stored overnight at 4° C. Following these treatments, 100 mL aliquots from each cell suspension were used to inoculate triplicate tubes of enriched growth medium (ESW; 0.2% glucose, 0.5% yeast extract, w/v in artificial seawater). Cultures were then incubated for three days at 25° C, with absorbance of cultures at 670 nm being measured via photospectrometry at 24-hour intervals. Triplicate culture tubes of ESW, not inoculated with bacteria, were incubated at 25° C as a negative control. Growth rates (Δ absorbance day^{-1}) were estimated for each experimental replicate by solving the exponential growth formula based on absorbance measurements after three days' incubation:

$$\frac{ln\,\frac{x_3}{x_0}}{3} = k$$

Where x_0 corresponds to the culture absorbance at t_0, x_3 corresponds to the culture absorbance at t_3, and k corresponds to the growth rate of the culture.

C.2. Prey viability tests results

Average growth rates for *P. bermudensis* cultures started from cell suspensions incubated at normal growth temperatures were 0.27 ± 0.02. Average growth rates for *P. bermudensis* cultures (started from cell suspensions inoculated with Cell Tracker, incubated three hours at 37° C, and then stored overnight at 4° C) were 0.32 ± 0.02. Average growth rates for *P. bermudensis*

cultures started from cell suspensions incubated three hours at 45° C were 0.09 ± 0.095. Growth rates for *P. bermudensis* cultures started from cell suspensions inoculated with Cell Tracker were significantly greater than those started from cell suspensions maintained at normal growth temperatures (Student's t-test, $p = 0.03$, $n = 3$). Growth rates for *P. bermudensis* cultures started from cell suspensions incubated at 45° C were not significantly different than growth rates for negative controls (Student's t-test, $p = 0.15$, $n = 3$).

C.3. Confocal microscopic examination of bacterivory by green algae

Confocal microscopy was conducted on the same treatments as for epifluorescence using a LSM710 (Carl Zeiss, Germany) microscope with an argon/2 laser to excite at 488 nm and optical sections were acquired at 517 nm and 680 nm for Cell Tracker Green or DTAF (FLB) and chlorophyll autofluorescence, respectively, to detect presence of ingested bacterial prey in the phagotrophic algal cells. Images were collected with a mounted digital microscope camera AxioCam.

C.4. Feeding experiments using magnetic beads

The chosen strains of green algae were also subjected to feeding experiments with magnetic beads. We selected magnetic beads for ingestion experiment with the original intention to use it for green algal phagosome isolation and proteomics characterization in a follow-up study (Lönnbro et al. 2008). Experiments were carried out with 0.5 µm Mono Mag Carboxylic Acid Beads (Ocean Nanotech, San Diego, California) or 1 µm Dynabeads MyOne Carboxylic Acid (ThermoFisher Scientific). Algae at different growth phases were pelleted, re-suspended in an Eppendorf tube, given sufficient time (2 days) to regrow flagella and then fed with magnetic beads. Prior to inoculation, the magnetic beads were washed 2× in filter-sterilized ASW and re-suspended in a thin-walled glass vial for 5-min sonication (Branson 200 Ultrasonic Cleaner) in

order to break apart bead aggregates. Beads were then added to the cells at a concentration of 1 µL beads per mL culture. Negative control samples received the same volume of sterile ASW. The obligate heterotrophic flagellate *Palpitomonas bilix* (Yabuki et al. 2010) was used as a positive control. After adding the beads or ASW, the mixture was incubated for 30 min on a hula mixer to stir the beads. In parallel, aliquots of the cultures were concentrated with the magnetic beads, to potentially increase encountering beads and thus increase consumption, by pelleting the cells after bead addition in a secondary Eppendorf at 500 g. Cells were observed under a Zeiss Axiovert 100M microscope, after 30-90 min since the start of incubation.

C.5. Detection limits and sources of variability in cytometry results

Detection of feeding in algal strains with low ingestion rates may have been limited by changes in fluorescence due to factors other than ingestion alone. Interpreting average per_{fed} values in PFA controls as a blank value, the ingestion detection limit three hours after inoculation with cell tracker is approximately 1.5 % (corresponding to the mean increase in per_{fed} observed in CT+PFA treatments plus two times the standard error). Additionally, activated Cell Tracker dye was commonly observed to accumulate in FLB suspension following the labeling protocol, presumably due to activation of dye by extracellular enzymes (First et al. 2012). The presence of activated dye was repeatedly observed to cause elevated background fluorescence both in cytometry and microscopy samples, even following the introduction of additional wash steps and the storage of FLB suspensions at 4 °C. While initial background fluorescence was accounted for in cytometry experiments by taking measurements of algal cells immediately following inoculation with FLB, interpretation of results obtained when working with CellTracker could be complicated due to photobleaching of extracellular fluorophores over the course of incubations. Especially in strains with low feeding rates, such as *D. tenuilepis*, increases in algal cell

fluorescence measured via flow cytometry may in fact reflect a net change, capturing the sum of 1) increases in fluorescence due to ingestion of labeled prey and 2) decreases in fluorescence due to photobleaching. This may explain the decrease in algal cell fluorescence over time in the case of *D. tenuilepis*, as well as the occasional observation of feeding in CT-FLB inoculated nutrient replete cultures via microscopy but not via flow cytometry. Although it is possible to eliminate effects of background fluorescence in cytometry experiments by reducing the sensitivity of the green fluorescence photomultiplier tubes, it is unclear what criteria should be used to determine the magnitude of this adjustment across different strains. Therefore, the same sensitivity settings were used for all experiments in this study.

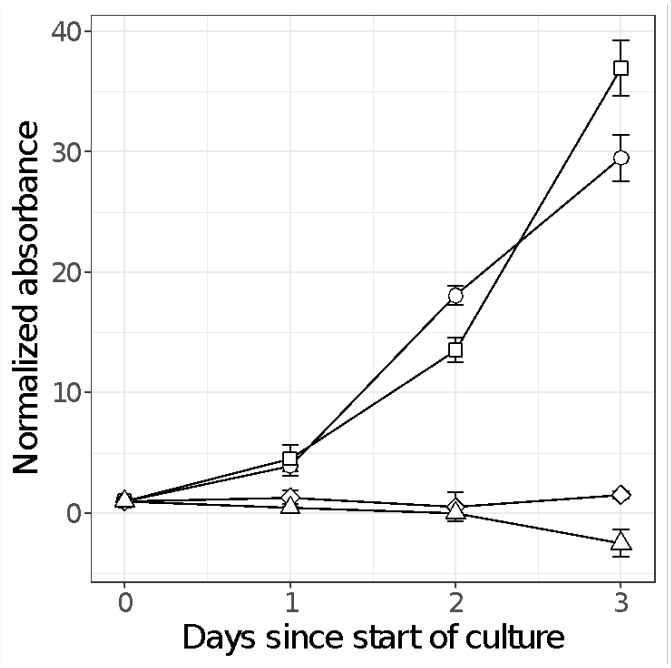

Figure C.1

Culture absorbance over three days in *Pelagibaca bermudensis* viability tests. Square markers correspond to cultures started from *P. bermudensis* cell suspensions following labeling protocol and overnight storage at 4° C; diamond markers correspond to cultures started from *P. bermudensis* cell suspensions incubated at 45° C; circle markers correspond to cultures started from *P. bermudensis* cell suspensions maintained under normal growth conditions; triangle markers correspond to absorbance of growth medium alone.

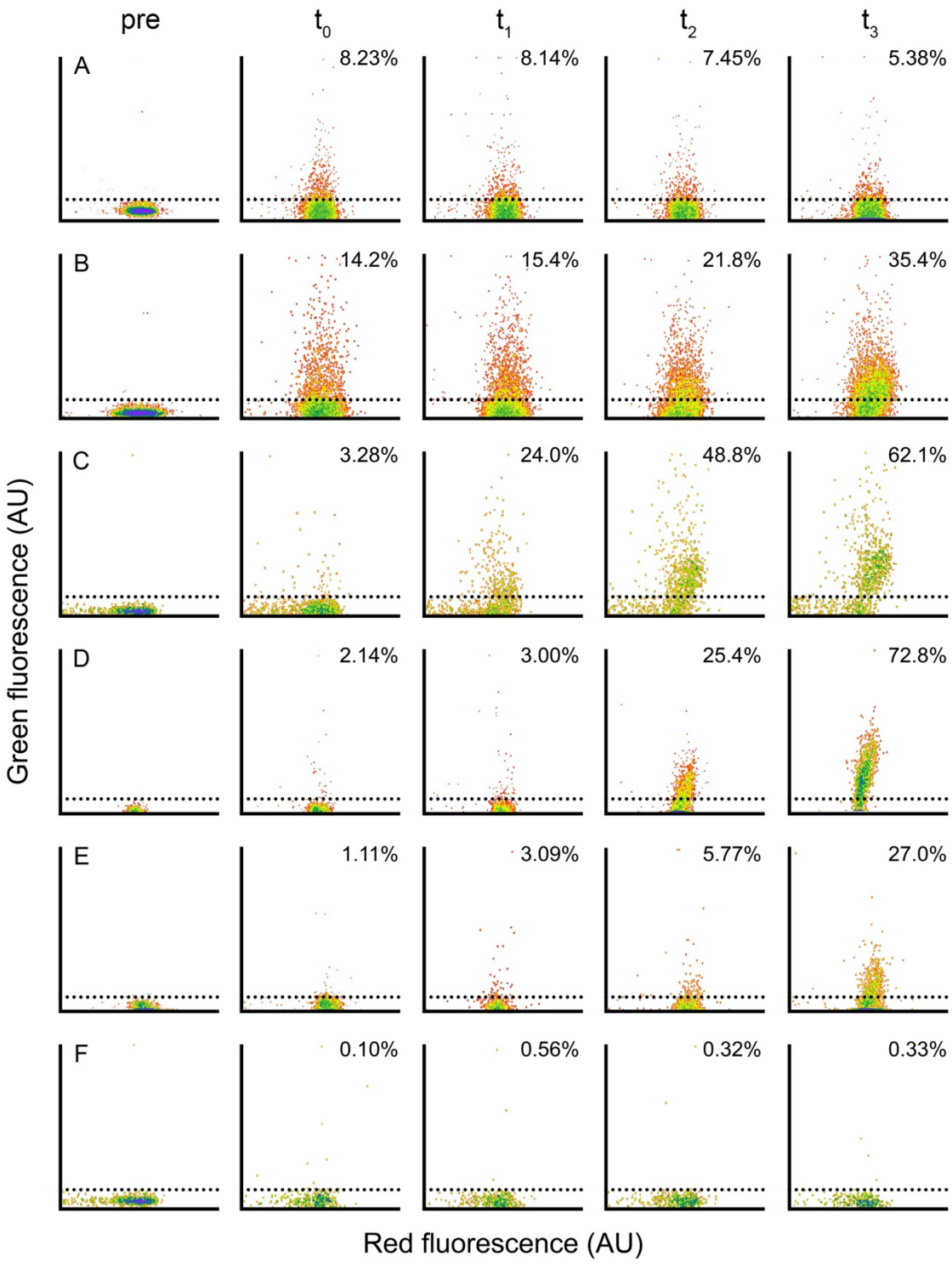

Figure C.2

Representative cytometry results for nutrient-limited treatments inoculated with CT-FLB for *Dolichomastix tenuilepis* (A), *Nephroselmis pyriformis* (B), *Pterosperma cristatum* (C),

Pyramimonas parkeae CCMP726 (D), and *Pyramimonas parkeae* NIES 254 (E). Representative cytometry results for nutrient-limited treatments inoculated with CT-FLB+PFA for *Pterosperma cristatum* (F). Cytograms correspond to cytometry measurements made prior to inoculation with CT-FLB (pre), or to cytometry measurements made 0 to 3 hours following inoculation with CT-FLB ($t_0 – t_3$). Dotted line corresponds to GF_{prior}. Percent values correspond to the percentage of total cells exceeding GF_{prior} at each time point. Events plotted based red fluorescence (arbitrary units; AU) and green fluorescence (AU). Cytograms were centered on the x axis and scaled by a constant scaling factor for clarity.

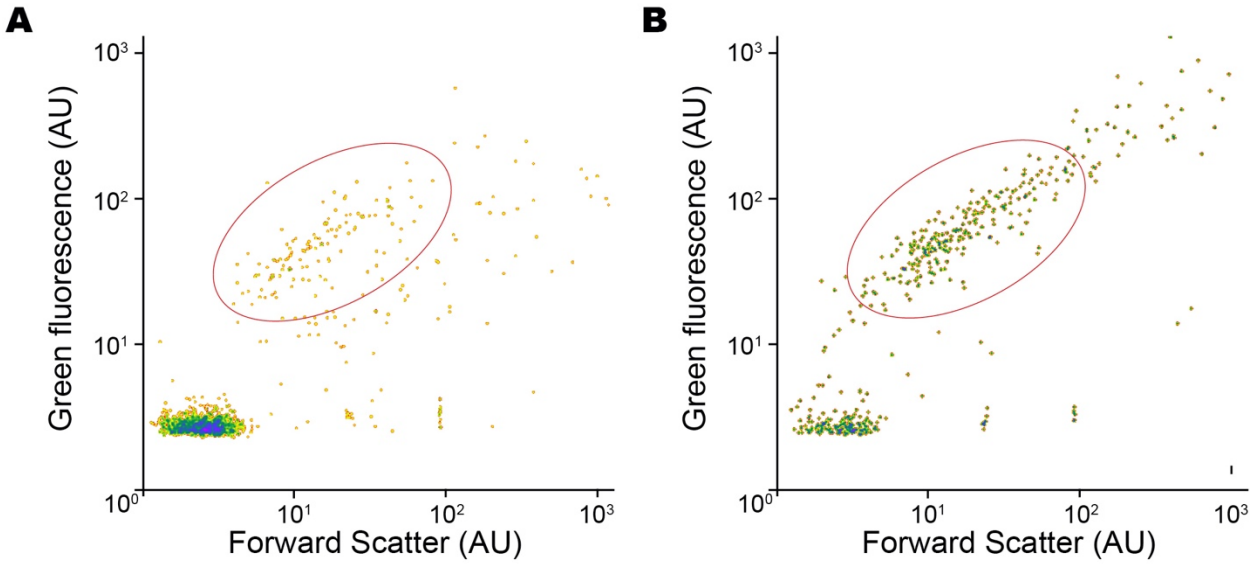

Figure C.3

Cytograms corresponding to representative dilutions of CT-FLB (A) and DTAF-FLB (B). Events plotted based on log-transformed forward scatter (arbitrary units; AU) and log-transformed green fluorescence. Red ellipses highlight populations of labeled bacteria. Events in lower left corner of cytograms correspond to electronic noise. Events not plotted below 2.5 AU on y axis due to green fluorescence detector threshold settings.

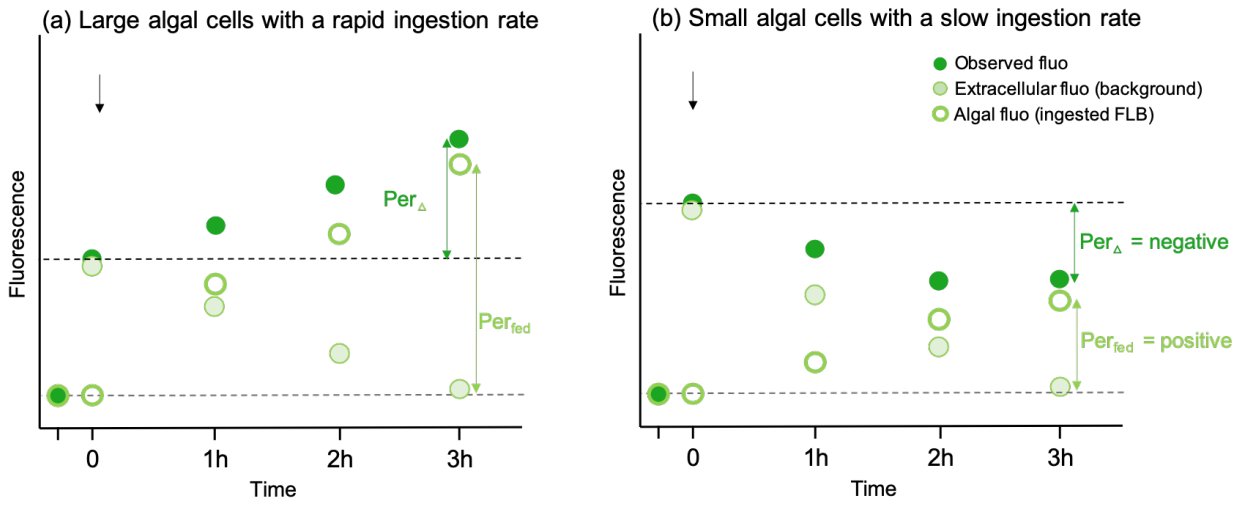

Figure C.4:

Conceptual representation of sources of underestimation in cytometry results of feeding experiments. The dark green dots represent the fluorescence observed with flow cytometry, which is the sum of the background fluorescence (light green dots) caused by extracellular fluorophores and the algal fluorescence (empty green circles) caused by ingestion of FLB. This conceptual model shows two contrasting trends in observed fluorescence over time as algal cells ingest the FLB and the extracellular fluorophores are bleached. For large algal cells (a), the bleaching of the initial background fluorescence leads to an underestimation of the percent feeding cells ($per_{\Delta} < per_{fed}$). For small cells (b) a greater initial background fluorescence and slower ingestion rates lead to an observed net decrease in fluorescence, masking the actual increase in algal fluorescence by ingestion ($per_{\Delta} < 0$; $per_{fed} > 0$).

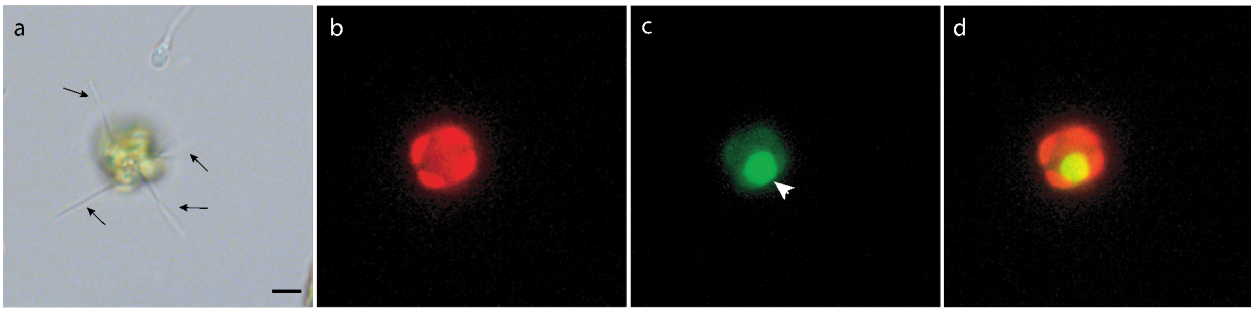

Figure C.5

Additional light microscopy images of Pyramimonas parkeae NIES254. The prasinophyte cells were fed with CT-FLB. From left to right: differential interference contrast (DIC)(a), chlorophyll fluorescence channel (b), FITC fluorescence channel (c), and an overlay of chlorophyll and FITC fluorescence images (d). Four flagella of the alga are indicated with arrows (a). The compartment where ingested CT-FLB accumulated is indicated with an arrowhead (C). Scale bar: 10 μm.

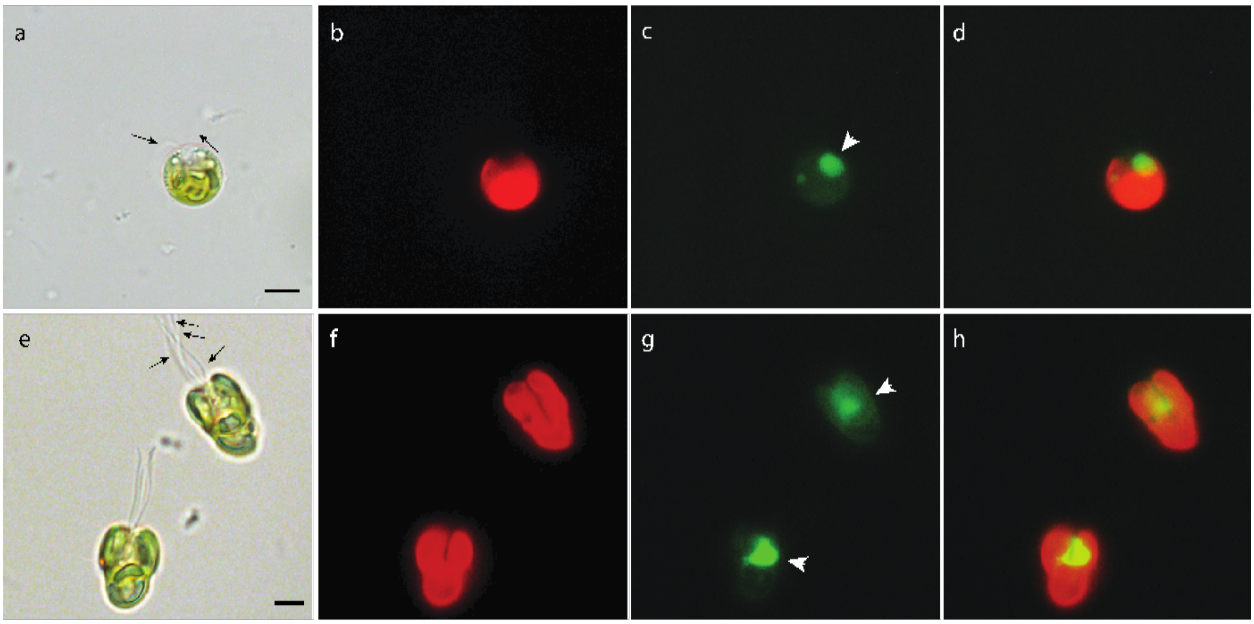

Figure C.6

Additional light microscopy images of *Pyramimonas parkeae* CCMP726. The prasinophyte cells were fed with CT-FLB. From left to right: differential interference contrast (DIC)(a, e), chlorophyll fluorescence channel (b, f), FITC fluorescence channel (c, g), and an overlay of chlorophyll and FITC fluorescence images (d, h). The flagella of the alga are indicated with arrows (a, e). The intracellular compartment where the ingested CT-FLB accumulated is indicated with an arrowhead (c, g). Scale bars: 10 µm. Note that the cell presented in the top panels is more rounded in the posterior end compared to those presented below. This difference in morphology is likely due to the size of starch storage, which accumulates in the cell's posterior.

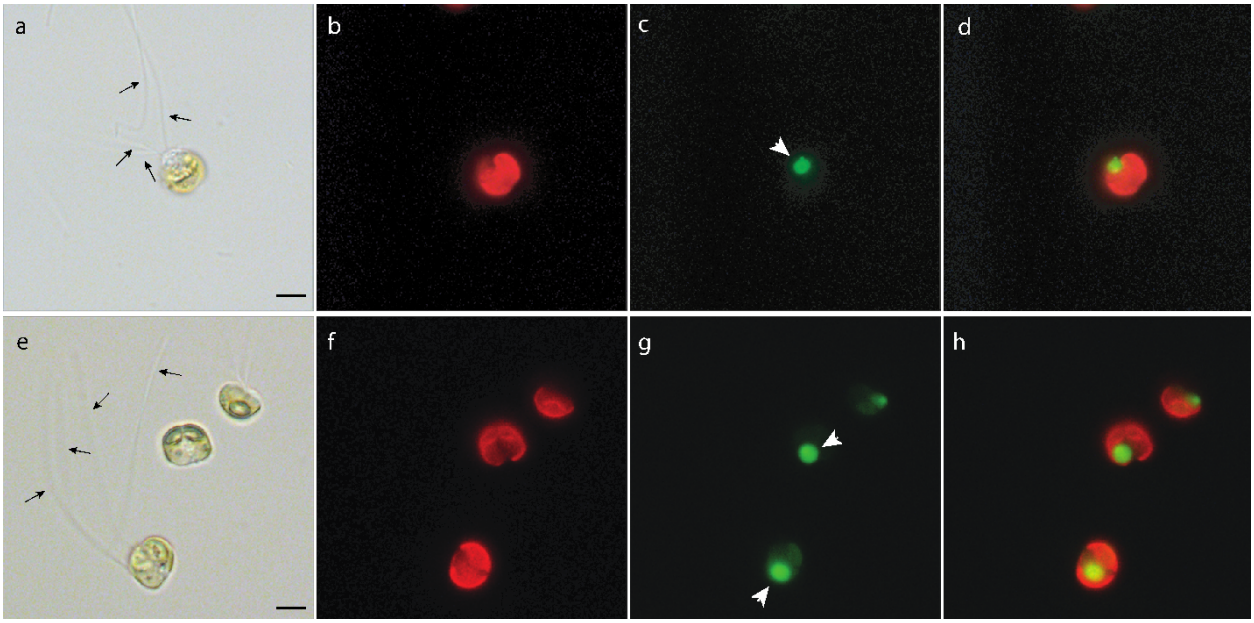

Figure C.7

Additional light microscopy images of Pterosperma cristatum NIES626. The prasinophyte cells were fed with CT-FLB. From left to right: differential interference contrast (DIC)(a, e), chlorophyll fluorescence channel (b, f), FITC fluorescence channel (c, g), and an overlay of chlorophyll and FITC fluorescence images (d, h). Four flagella of the alga are indicated with arrows (a, e). The spherical compartment where the ingested CT-FLB accumulated is indicated with an arrowhead (c, g). Scale bars: 10 µm.

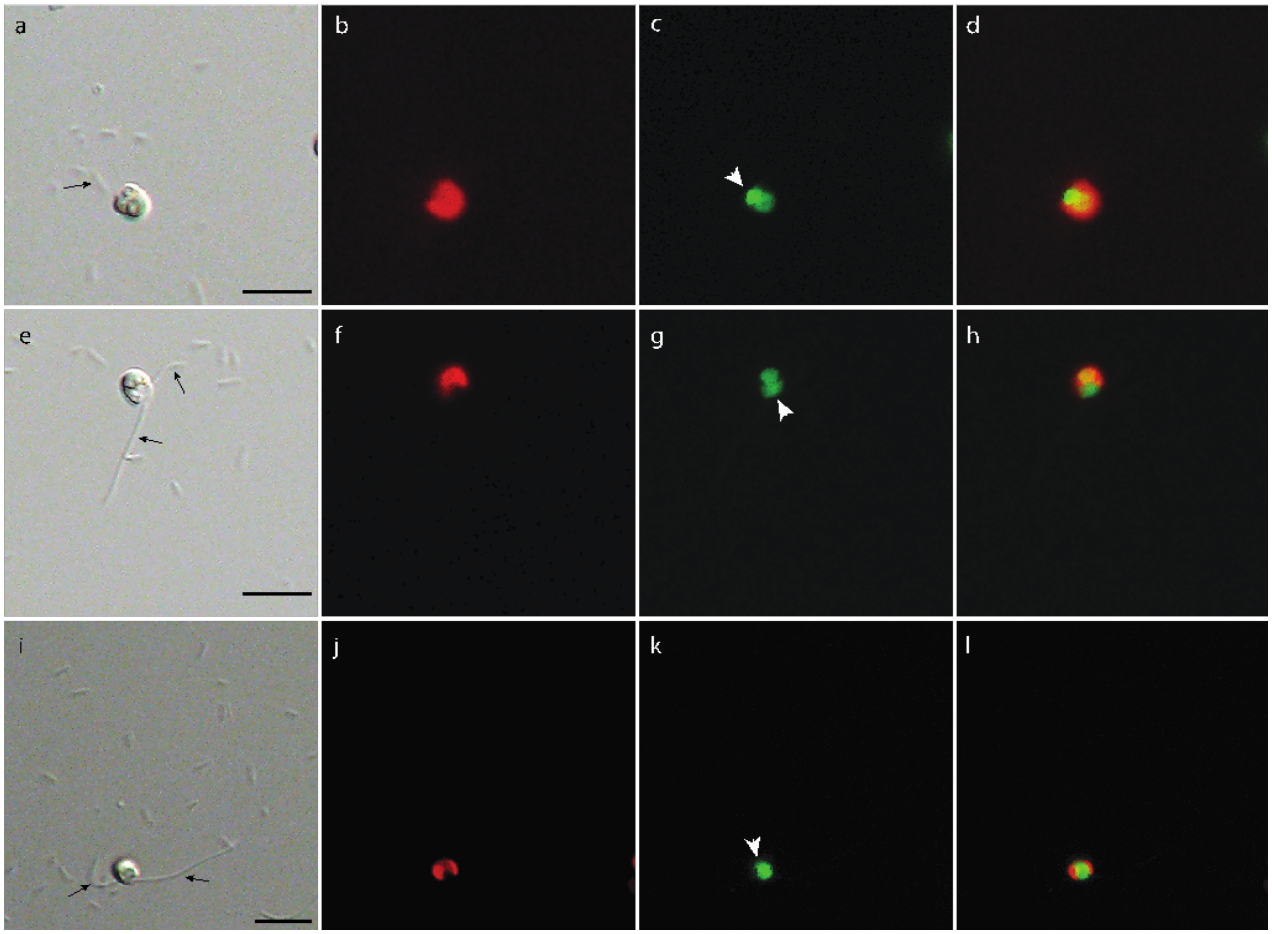

Figure C.8

Additional light microscopy images of *Nephroselmis pyriformis* RCC618. The prasinophyte cells were fed with CT-FLB. From left to right: differential interference contrast (DIC)(a, e, i), chlorophyll fluorescence channel (b, f, j), FITC fluorescence channel (c, g, k), and an overlay of chlorophyll and FITC fluorescence images (d, h, l). The alga bears two flagella, indicated with arrows (a, e, i). The intracellular compartment where the ingested CT-FLB accumulated is indicated with an arrowhead (c, g, k). Scale bars: 10 µm.

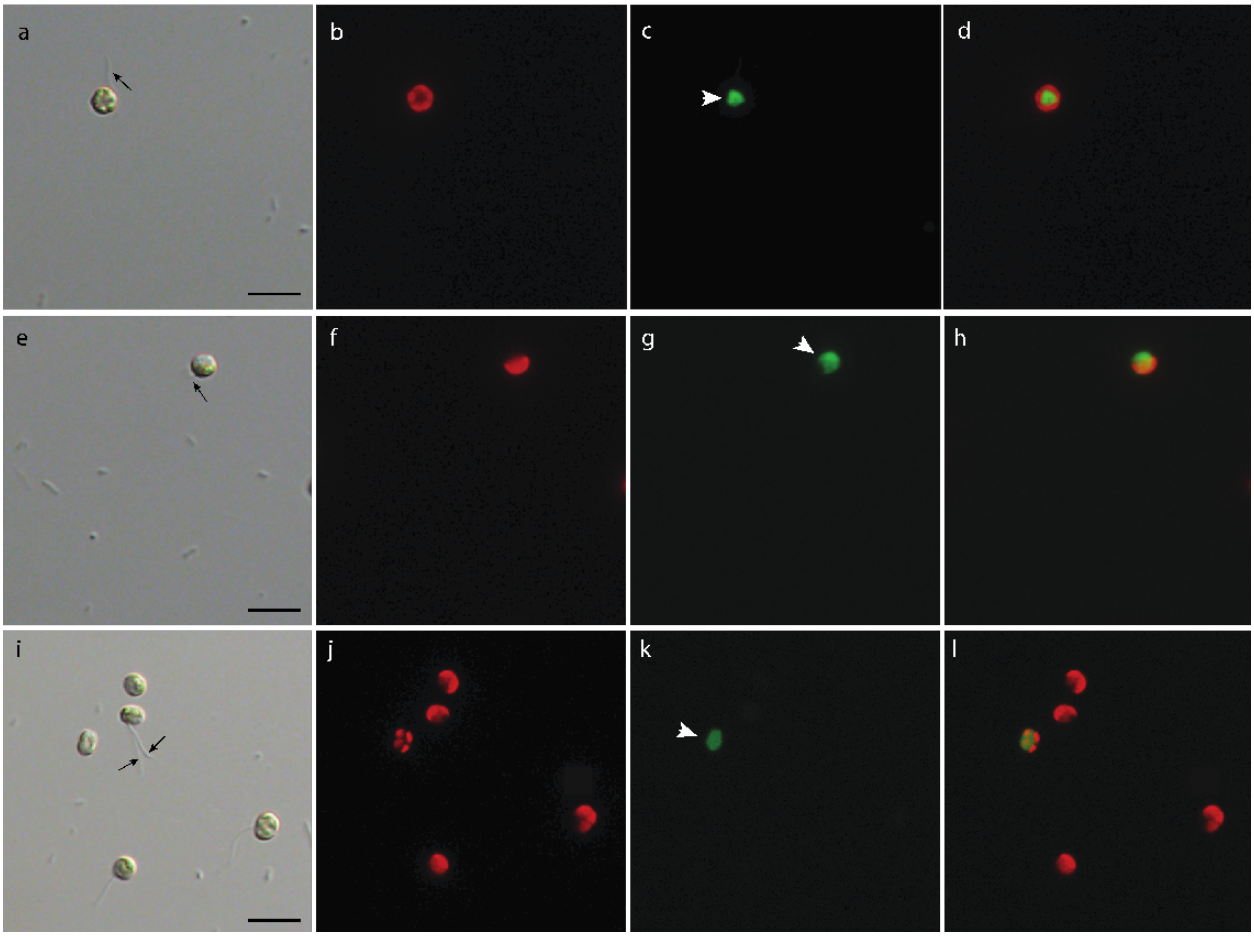

Figure C.9

Additional light microscopy images of *Dolichomastix tenuilepis* CCMP3274. From left to right: differential interference contrast (DIC)(a, e, i), chlorophyll fluorescence channel (b, f, j), FITC fluorescence channel (c, g, k), and an overlay of chlorophyll and FITC fluorescence images (d, h, l). The alga bears two flagella, indicated with arrows (a, e, i). The intracellular compartment where the ingested CT-FLB accumulated is indicated with an arrowhead (c, g, k). Scale bars: 10 µm.

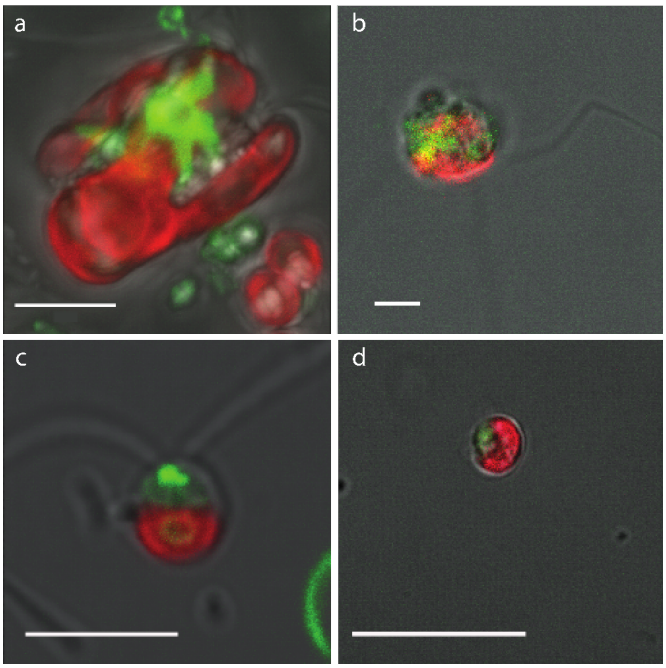

Figure C.10

Images from confocal microscopy of *Pyramimonas parkeae* NIES254 (a), *Pterosperma cristatum* NIES626 (b), *Nephroselmis pyriformis* RCC618 (c) and *Dolichomastix tenuilepis* CCMP3274 (d) fed with a CT-labeled *Pelagibaca bermudensis* HTCC2601. Images acquired at 517 nm (for detecting the labeled bacteria) and 680 nm (for detecting the chloroplast compartment) were overlaid onto the corresponding bright-field image. Scale bars: 10 µm in (a) and (b), 5 µm in (c) and (d).